

# Multiresolution Subband Blind Source Separation: Models and Methods

Hongwei Li<sup>1</sup>, Rui Li<sup>2</sup>, Fasong Wang<sup>1</sup>

1.School of Mathematics and Physics, China University of Geosciences, Wuhan 430074, P.R.China

Email: fasongwang@126.com, [hwli@cug.edu.cn](mailto:hwli@cug.edu.cn)

2.School of Sciences, Henan University of Technology, Zhengzhou 450052, P.R.China

Email: liruiyric@126.com

**Abstract**—The goal of this paper is to review the multiresolution subband Blind Source Separation (MRSBSS) methods, classify them into different categories, and identify new trends. After the perspective of MRSBSS is briefly presented, the general and detailed definition of the MSBSS model is given, the relationships among BSS, Independent Component Analysis (ICA) and MRSBSS method are discussed simultaneously. Then, the state-of-the-art MRSBSS algorithms: adaptive filter based method and sparsity based multiscale ICA method are constructed and overviewed in detail from different theory base. The paper concludes by discussing the impact of multiresolution subband on BSS research and outlining potential future research directions and applications.

**Index Terms**—Independent Component Analysis(ICA), Blind Source Separation(BSS), Independent Subspace Analysis(ISA), Subband; Multiresolution, Wavelet; Wavelet Packets

## I. INTRODUCTION

Standard Blind Source Separation (BSS) model and methods have been successfully applied to many areas of science [1-2]. The basic BSS model assumes that the observed signals are linear superposition's of underlying hidden source signals. Most of the BSS algorithms are based on the independent assumption of the source signals which is called Independent Component Analysis (ICA). Despite the success of using standard ICA in many applications, the basic assumptions of ICA may not hold hence some caution should be taken when using standard ICA to analyze real world problems, especially in biomedical signal processing and image processing. In fact, by definition, the standard ICA algorithms are not able to estimate statistically dependent original sources, that is, when the independence assumption is violated.

Among many extensions of the basic ICA model, 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 DCA model is the Multidimensional Independent Component Analysis (MICA) model[3], which is a linear generative model as ICA. 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], topographic ICA[6], and tree-dependent component analysis[7], subband decomposition ICA[8], maximum non-Gaussianity method [9-10], spectral decomposition method[11], time-frequency method[12-13].

In these methods, the subband decomposition ICA[8] method can be naturally extended and generalized to multiresolution subband BSS(MRSBSS) which relaxes considerably the assumption regarding mutual independence of primarily sources. The more powerful assumption is non-independent of the sources decomposition coefficients, when the signals are properly represented or transformed. Multiresolution analysis (MRA) will show how discrete signals are synthesized by beginning with a very low resolution signal and successively adding on details to create higher resolution versions, ending with a complete synthesis of the signal at the finest resolution. This is known as. In these transformed fields, the coefficients do not necessarily have to be linearly independent and, instead, may form an overcomplete set (or dictionary), for example, wavelet packets, stationary wavelets, time-frequency transformations. The key idea in this approach is the assumption that the wide-band source signals or their transformations are dependent, however some narrow band subcomponents are independent. In other words, we assume that each unknown source can be modeled or represented as a sum of narrow-band sub-signals (sub-components)[2]:

$$s_i(t) = s_{i,1}(t) + s_{i,2}(t) + \dots + s_{i,K}(t) \quad (1.1)$$

The basic concept of MRSBSS is to divide the sensor signal spectra into their subspectra or subbands, and then to treat those subspectra or subbands individually for the

purpose at hand. The subband signals can be ranked and processed independently. Let us assume that only a certain set of sub-components is independent. Provided that for some of the transformation subbands (at least one) of all sub-components, say  $s_{ip}, i=1, \dots, N; p \in \{1, \dots, K\}$ , are mutually independent or temporally decorrelated, then we can easily estimate the mixing or separating system (under condition that these subbands can be identified by some a priori knowledge or detected by some self-adaptive process) by simply applying any standard BSS/ICA algorithm, however not for all available raw sensor data but only for suitably preprocessed (band pass filtered) sensor signals. Such explanation can be summarized as follows.

In one of the most simplest cases, source signals can be modeled or decomposed into their low- and high-frequency sub-components:

$$s_i(t) = s_{i,H}(t) + s_{i,L}(t) \tag{1.2}$$

In practice, the high-frequency sub-components  $s_{i,H}(t)$  are often found to be mutually independent. In such a case in order to separate the original sources  $s_i(t)$ , we can use a High Pass Filter (HPF) to extract high frequency sub-components and then apply any standard ICA algorithm to such preprocessed sensor (observed) signals. In the preprocessing stage, more sophisticated methods, such as block transforms, multi-rate subband filter bank or wavelet transforms, can be applied.

**Definition 1.6<sup>[2]</sup> (Multiresolution Subband Blind Signal Separation: MRSBSS)** *The MRSBSS can be formulated as a task of estimation of the mixing matrix on the basis of suitable multiresolution subband decomposition of sensors signals and by applying a classical BSS method (instead for raw sensor data) for one or several pre-selected subbands for which source sub-components are least dependent.*

## II. MULTIREOLUTION SUBBAND BSS MODEL

### A. BSS model

Let us denote the  $N$  source signals by the vector  $\mathbf{s}(t) = (s_1(t), \dots, s_N(t))^T$ , and the observed signals by  $\mathbf{x}(t) = (x_1(t), \dots, x_M(t))^T$ . Now the mixing can be expressed as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \tag{2.1}$$

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.  $\mathbf{n}(t) = (n_1(t), \dots, n_M(t))^T$  is a vector of additive noise. 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].

The task of BSS 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,  $\mathbf{y}(t) = (y_1(t), \dots, y_n(t))^T$  is an estimate of the possibly scaled and permuted 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  $\mathbf{s}(t)$ .

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  $y_i(t)$ ,  $i = 1, 2, \dots, n$  are mutually independent too [14].

*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)  $\mathbf{C} = \mathbf{Q}\mathbf{A}$ , where  $\mathbf{Q}$  is a permutation matrix and  $\mathbf{A}$

is a nonsingular diagonal matrix.

To cope with ill-conditioned cases and to make algorithms simpler and faster, a linear transformation called prewhitening is often used to transform the observed signals  $\mathbf{x}$  to  $\tilde{\mathbf{x}} = \mathbf{V}\mathbf{x}$  such that

$$E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] = \mathbf{I},$$

where  $\mathbf{V}$  is a prewhitening matrix.

*B MRSBSS Model*

The assumption of statistical independence of the source components  $s_i(t)$  leads itself to ICA and justified by the physics of many practical applications. A more powerful assumption is non-independent of the decomposition coefficients when the signals are properly represented[15-18]. Suppose that each  $s_i(t)$  have a transformed representation of its decomposition coefficients  $c_{ik}$  obtained by means of the set of representation functions  $\{\varphi_k(t), k = 1, \dots, K\}$ :

$$s_i(t) = \sum_{k=1}^K c_{ik} \varphi_k(t) \tag{2.2}$$

The functions  $\varphi_k(t)$  are called *atoms* or *elements* of the representation space that may constitute a basis or a framework. These elements  $\{\varphi_k(t), k = 1, \dots, K\}$  do not have to be linearly independent and, instead, may be mutually dependent and form an overcomplete set (or dictionary), for example, wavelet-related dictionaries: wavelet packets, stationary wavelets, etc. [19-20]. The corresponding representation of the mixtures  $\mathbf{x}(t)$ , according to the same signal dictionary, is:

$$\mathbf{x}_i(t) = \sum_{k=1}^K d_{ik} \varphi_k(t) \tag{2.3}$$

$$\mathbf{n}_i(t) = \sum_{k=1}^K e_{ik} \varphi_k(t) \tag{2.4}$$

where  $d_{ik}$  and  $e_{ik}$  are the decomposition coefficients of the *mixtures and noises*.

Let us consider the case wherein the subset of functions  $\{\varphi_k(t), k \in \Omega\}$  is constructed from the mutually orthogonal elements of the dictionary. This subset can be either complete or undercomplete.

Define vectors  $\{\mathbf{d}_l, l \in \Omega\}$ ,  $\{\mathbf{c}_l, l \in \Omega\}$  and  $\{\mathbf{e}_l, l \in \Omega\}$ . to be constructed from the  $k$ -th coefficients of mixtures, sources and noises, respectively. From (2.1), (2.2)-(2.4), and using the orthogonality of the subset of functions  $\{\varphi_k(t), k \in \Omega\}$ , the relation between the decomposition coefficients of the mixtures and of the sources is

$$\mathbf{d}_l = \mathbf{A} \mathbf{c}_l + \mathbf{n}_l \tag{2.5}$$

Note, that the relation between decomposition coefficients of the mixtures and the sources is exactly the same relation as in the original domain of signals. Then, estimation of the mixing matrix and of sources is performed using the decomposition coefficients of the mixtures  $\mathbf{d}_l$  instead of the mixtures  $\mathbf{x}(t)$ .

The property of coefficients in the transform domain often yields much better source separation than standard BSS methods, and can work well even with more sources than mixtures. In many cases there are distinct groups of

coefficients, wherein sources have different statistical properties. The proposed multiresolution, or multiscale, approach to the BSS is based on selecting only a subband or subset of features, or coefficients,  $\mathbf{D} = \{\mathbf{d}_l, l \in \Omega\}$ , which is best suited for separation, with respect to the coefficients and to the separability of sources' features.

So, for MRSBSS model, in order to estimating the source signals, it is very important to choose the form of the transformation and the approach to select the optimal subband or subset of the coefficients. In the next section, we will review some algorithms to solve the MRSBSS model using different transformations and subband selection methods.

III. MRSBSS ALGORITHMS

Different approaches to transformations of the source signals and various methods to choose the optimal subset of the coefficients can result in a set of MRSBSS algorithms. The whole procedure of the MRSBSS, from mixing to separating, is described in figure 1. In this section, we will give an overview to these approaches.

*A Adaptive Filter Based Method*

Subband decomposition ICA(SDICA) [8,21-23], 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 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 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 [21], 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 [21], 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.

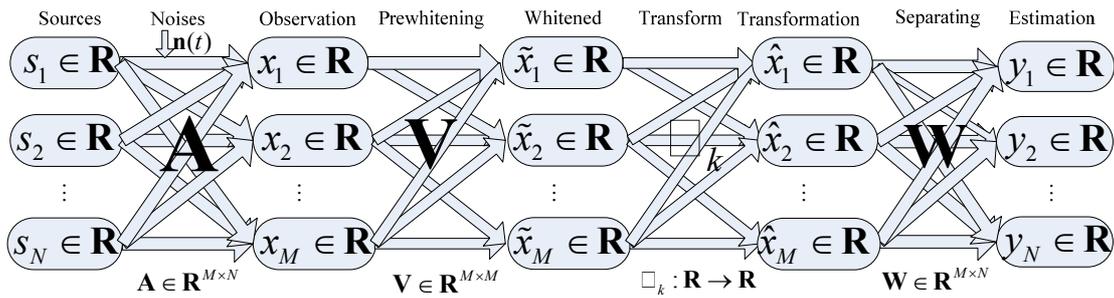


Figure 1. The whole mixing-separation procedure of MRSBSS, where  $\mathbf{s} = [s_1, \dots, s_N]^T \in \mathbf{R}^N$  is the hidden source signals;  $\mathbf{x} = [x_1, \dots, x_M]^T \in \mathbf{R}^M$  is the observation and  $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ ;  $\tilde{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_M]^T \in \mathbf{R}^M$  is the whitened mixed signals and  $\tilde{\mathbf{x}}(t) = \mathbf{V}\mathbf{x}(t)$ ;  $\hat{\mathbf{x}} = [\hat{x}_1, \dots, \hat{x}_M]^T \in \mathbf{R}^M$  is the transformed mixed signals and  $\hat{\mathbf{x}}(t) = \mathbf{T}_k[\tilde{\mathbf{x}}(t)]$ ;  $\mathbf{y} = [y_1, \dots, y_N]^T \in \mathbf{R}^N$  is the estimated sources and  $\mathbf{y}(t) = \mathbf{W}\hat{\mathbf{x}}(t)$ .

Zhang et.al. discuss the validity of the existing methods for SDICA and the conditions for separability of the SDICA model [8]. They propose an adaptive method for SDICA, called band-selective ICA (BS-ICA).

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 [21]. In other words, all sources  $s_i$  are not necessarily independent, but can be represented as the sum of several subcomponents as

$$s_i(t) = s_{i,1}(t) + s_{i,2}(t) + \dots + s_{i,L}(t) \tag{3.1}$$

where  $s_{i,k}(t)$ ,  $k=1,2,\dots,L$  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 (2.1). 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.2. (a). 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  $\mathbf{s}^{(k)}(t) = [s_{1,k}(t), \dots, s_{n,k}(t)]^T$ . 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  $\mathbf{A}$ .

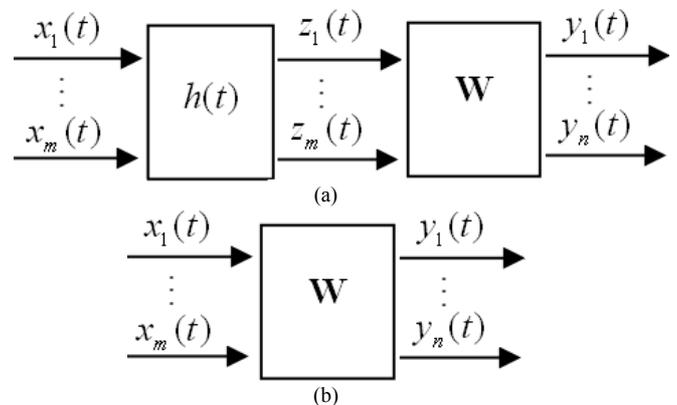


Figure 2. (a) The structure to perform SDICA. (b) The linear instantaneous ICA demixing.

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  $\mathbf{A}$  and the source-independent subcomponents can be recovered. Band-selective ICA (BS-ICA) is such a method [11].

In [8], authors 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, [8] apply a linear filter on each observation, 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 details can be found in [8].

**Adjusting  $\mathbf{W}$**  [8]

$$\begin{aligned} \mathbf{W}(k+1) - \mathbf{W}(k) = & \eta(k) [\mathbf{I} - \text{diag}\{E(y_1^2), \dots, E(y_N^2)\} \\ & + E(\varphi_y(\mathbf{y}(k))\mathbf{y}(k)^T) \\ & - \text{diag}\{E[\varphi_y(\mathbf{y}(k))\mathbf{y}(k)^T]\}] \mathbf{W}(k), \end{aligned} \tag{3.3}$$

where  $\varphi_y(\mathbf{y}) = [\varphi_{y_1}(y_1), \dots, \varphi_{y_N}(y_N)]^T$  are the so-called score functions.,  $\text{diag}\{E(y_1^2), \dots, E(y_N^2)\}$  denotes the diagonal matrix with  $E(y_1^2), \dots, E(y_N^2)$  as its diagonal entries, and  $\text{diag}\{E[\varphi_y(\mathbf{y})\mathbf{y}^T]\}$  denotes the diagonal matrix whose diagonal entries are those on the diagonal of  $E[\varphi_y(\mathbf{y})\mathbf{y}^T]$ .

**Adjusting  $h(t)$**  [8]

$$\begin{aligned} \frac{\partial I(\mathbf{y})}{\partial h_k} = & \sum_{i=1}^N \frac{\partial H(y_i)}{\partial h_k} - \frac{\partial H(\mathbf{y})}{\partial h_k} \\ = & -E_t\{\varphi_y^T(t) \cdot \mathbf{W} \cdot \mathbf{x}(t-k)\} \\ & + E_t\{\psi_y^T(t) \cdot \mathbf{W} \cdot \mathbf{x}(t-k)\} \\ = & E_t\{\beta_y^T(t) \cdot \mathbf{W} \cdot \mathbf{x}(t-k)\} \end{aligned} \tag{3.4}$$

where  $\beta_y^T(t) = \varphi_y(\mathbf{y}) - \psi_y(\mathbf{y})$  is defined as the score function difference (SFD). The SFD is an independence criterion; it vanishes if and only if  $y_i$  are mutually independent. Now the elements of  $h(t)$  can be adjusted with the gradient-descent method.

Eq.(3.3) and eq.(3.4) are the learning rules of BS-ICA. When the instantaneous stage and the filter stage both converge, the matrix  $\mathbf{W}$  is the demixing matrix associated with the mixing matrix  $\mathbf{A}$ . The original sources can be estimated as  $\mathbf{W}\mathbf{x}$ . Moreover, the outputs  $y_i$  form the estimate of a filtered version of the independent subcomponents  $s_{i,l}$ .

*B Sparsity Based Multiresolution ICA Method*

As the basic BSS model, assumptions can be made about the nature of the sources. Such assumptions form the basis for most source separation algorithms and include statistical properties such as independence and stationarity. One 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. A signal is said to be sparse when it is zero or nearly zero more than might be expected from its variance. Such a signal has a probability density function or distribution of values with

a sharp peak at zero and fat tails. This shape can be contrasted with a Gaussian distribution, which would have a smaller peak and tails that taper quite rapidly. A standard sparse distribution is the Laplacian distribution, which has led to the sparseness assumption being sometimes referred to as a Laplacian prior. The advantage of a sparse signal representation is that the probability of two or more sources being simultaneously active is low. Thus, sparse representations lend themselves to good separability because most of the energy in a basis coefficient at any time instant belongs to a single source. This statistical property of the sources results in a nicely defined structure being imposed by the mixing process on the resultant mixtures, which can be exploited to make estimating the mixing process much easier. Additionally, sparsity can be used in many instances to perform source separation in the case when there are more sources than sensors. A sparse representation of an acoustic signal can often be achieved by a transformation into a Fourier, Gabor or Wavelet basis[26].

Sparse sources can be separated by each one of several techniques, for example, by approaches based on the maximum likelihood (ML) considerations[27], or by approaches based on geometric considerations[28]. In the former case, the algorithm estimates the unmixing matrix  $\mathbf{W} = \mathbf{A}^{-1}$ , while in the later case the output is the estimated mixing matrix [29]. Here, we only consider the former case.

Now, we discuss the ML and the quasi<sup>1</sup> ML solution of the BSS problem, based on the data in the domain of decomposition coefficients.

Let  $\mathbf{D}$  be the new coefficient matrix of dimension  $N \times T$ , or called features, where  $T$  is the number of data points, and the coefficients  $\mathbf{d}_i$  form the columns of  $\mathbf{D}$ . Note that, in general, the rows of  $\mathbf{D}$  can be formed from either the samples of mixtures, or from their decomposition coefficients. In the latter case,  $\mathbf{D} = \{\mathbf{d}_i : i \in \Omega\}$ , where  $\Omega$  are the subsets indexed on the transformation. We are interested in the maximum likelihood estimate of  $\mathbf{A}$  given the data  $\mathbf{D}$ .

It is assumed that that the coefficients of the source signals  $c_{ik}$  are i.i.d. random variables with the joint probability density function (pdf)

$$p(\mathbf{C}) = \prod_{i,k} p(c_{ik})$$

where  $p(c_{ik})$  is of an exponential type:

$$p(c_{ik}) = N_q \exp\{-\nu(c_{ik}, q)\}$$

where  $N_q$  is the normalization constant which will be omitted in the further calculations, since it has no effect on the maximization of the log-likelihood function. In a particular case wherein

<sup>1</sup> The term quasi indicates that the proposed hypothetical density is used instead of the true one.

$$v(c_{ik}, q) = \frac{|c_{ik}|^q}{q} \tag{3.5}$$

and  $q < 1$ , the above distribution is widely used for modeling sparsity. For  $q = 1/2$  it approximates rather well the empirical distributions of wavelet coefficients of natural signals and images[30]. A smaller  $q$  corresponds to a distribution with greater super-Gaussian that is more sparsity.

Let  $\mathbf{W}^o \mathbf{A}^{-1}$  be the unmixing matrix to be estimated. Taking into account that

$$\mathbf{D} = \mathbf{A}\mathbf{C} + \mathbf{N},$$

If we neglect the effect of the noise, we arrive at the standard expression of the ICA log-likelihood, but with respect to the decomposition coefficients:

$$L(\mathbf{D}, \mathbf{W}) = \log p(\mathbf{D}) = K \log |\det(\mathbf{W})| - \sum_{i=1}^M \sum_{k=1}^K v([\mathbf{W}\mathbf{D}]_{ik}, q) \tag{3.6}$$

In the case wherein (3.5), the second term in the above log-likelihood function (3.6) is not convex for  $q < 1$ , and non-differentiable, therefore it is difficult to optimize. Furthermore, the parameter  $q$  of the true pdf is usually unknown in practice, and estimation of this parameter along with the estimation of the unmixing matrix is a difficult optimization problem. Therefore, it is convenient to replace the actual  $v(\square)$  with its hypothetical substitute, a smooth, convex approximation of the absolute value function, for example  $\mathcal{V}(c_{ik}) = \sqrt{c_{ik}^2 + z}$ , with  $z$  being a smoothing parameter. This approximation has a minor effect on the separation performance, as indicated by our numerical results. The corresponding *quasi* log-likelihood function is

$$\tilde{L}(\mathbf{D}, \mathbf{W}) = K \log |\det(\mathbf{W})| - \sum_{i=1}^M \sum_{k=1}^K \tilde{v}([\mathbf{W}\mathbf{D}]_{ik}) \tag{3.7}$$

Maximization of  $\tilde{L}(\mathbf{D}, \mathbf{W})$  with respect to  $\mathbf{W}$  can be solved efficiently by several methods, for example, the Natural Gradient (NG) algorithm [31] or, equivalently, by the Relative Gradient [32], as implemented in the BSS Matlab toolbox [33].

The derivative of the quasi log-likelihood function with respect to the matrix parameter  $\mathbf{W}$  is:

$$\frac{\partial \tilde{L}(\mathbf{D}, \mathbf{W}(k))}{\partial \mathbf{W}(k)} = [\mathbf{K}\mathbf{I} - \varphi(\mathbf{c}(k))\mathbf{c}(k)^T] (\mathbf{W}(k)^T)^{-1} \tag{3.8}$$

where  $\mathbf{c}$  is the 'column stack' version of  $\mathbf{C}$ , and

$$\varphi(\mathbf{c}) = [\varphi(c_1), \dots, \varphi(c_{NK})]^T,$$

where  $\varphi_{ik}(c_{ik}) = -\frac{d}{dc_{ik}} \log p_{ik}(c_{ik}) = -\frac{p'_{ik}(c_{ik})}{p_{ik}(c_{ik})}$  are the so-called score functions. Note, that, in this case,

$\varphi_{ik}(c_{ik}) = \tilde{v}'(c_{ik})$ . The learning rule of the Natural Gradient algorithm is given by

$$\begin{aligned} \mathbf{W}(k+1) - \mathbf{W}(k) &= \frac{\partial \tilde{L}(\mathbf{D}, \mathbf{W}(k))}{\partial \mathbf{W}(k)} (\mathbf{W}(k))^T \mathbf{W}(k) \\ &= \eta(k) [\mathbf{K}\mathbf{I} - \varphi(\mathbf{c}(k))\mathbf{c}(k)^T] \mathbf{W}(k), \end{aligned} \tag{3.9}$$

where  $\eta(k) > 0$  is the learning rate. It has been proven that the natural gradient greatly improves learning efficiently in ICA than gradient method. In the original Natural Gradient algorithm, the above update equation is expressed in terms of the non-linearity, which has the form of the cumulative density function (cdf) of the hypothetical source distribution.

In order to adapt the algorithm to purposes, we use hypothetical density with  $\mathcal{V}(c_{ik}) = \sqrt{c_{ik}^2 + z}$  and the corresponding score function is [29]

$$\varphi(c_{ik}) = \frac{c_{ik}}{\sqrt{c_{ik}^2 + z}}$$

The estimated unmixing matrix  $\hat{\mathbf{W}} = \hat{\mathbf{A}}^{-1}$  can be obtained by the quasi-ML *approach, implemented along with the Natural Gradient*, or by other algorithms, which are applied to either the complete data set, or to some subsets of data. In any case, this matrix and, therefore, the sources, can be estimated only up to a column permutation and a scaling factor. The sources are recovered in their original domain by

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

The next question is how to choose the proper transformation to dispose the mixtures. In this paper, Wavelet Packet (WP) transform is used to reveal the structure of signals, wherein several subsets of the WP coefficients have significantly better sparsity and separability than others.

In particular case of WP, we express each source signal and noise in terms of its decomposition coefficients:

$$s_{ki}^j(t) = \sum_l c_{kil}^j \varphi_{jl}(t) \tag{3.10}$$

$$n_{ki}^j(t) = \sum_l e_{kil}^j \varphi_{jl}(t) \tag{3.11}$$

where  $j$  represents scale level,  $k$  represents sub-band index,  $i$  represents source index and  $l$  represents shift index.  $\varphi_{jl}(t)$  is the chosen wavelet also called atom or element of the representation space and  $c_{kil}^j$  are decomposition coefficients. In our implementation of the described SDICA algorithm, we have used 1D WP for separation the mixed signals. If we choose the same representation space as for the source signals, we express each component of the observed data  $\mathbf{x}$  as

$$x_{ki}^j(t) = \sum_l d_{kil}^j \varphi_{jl}(t) \tag{3.12}$$

where  $i$  represents sensor index.

Let vectors  $\mathbf{d}_l$ ,  $\mathbf{c}_l$  and  $\mathbf{e}_l$  be the constructed from the  $l$  th coefficients of the mixtures, sources and noises, respectively. From (2.1) and (3.12) using the orthogonality property of the functions  $\varphi_{jl}(t)$  we obtain

$$\mathbf{d}_l = \mathbf{A}\mathbf{c}_l + \mathbf{n}_l, \quad (3.13)$$

Thus, when the noise is present Eq. (15) becomes

$$\mathbf{d}_l \square \mathbf{A}\mathbf{c}_l. \quad (3.14)$$

Note, that the relation between decomposition coefficients of the mixtures and the sources is exactly the same relation as in the original domain of signals. Then, estimation of the mixing matrix and of sources is performed using the decomposition coefficients of the mixtures  $\mathbf{d}_l$  instead of the mixtures  $\mathbf{x}$ .

From (2.1) and (3.14) we see the same relation between signals in the original domain and WP representation domain. Inserting (3.14) into (3.12) and using (3.11) we obtain:

$$\mathbf{x}_k^j(t) = \mathbf{A}\mathbf{s}_k^j(t) \quad (3.15)$$

as it was announced by (3.8) where no assumption of multiscale decomposition has been made. For each component  $x_i$  of the observed data vector  $\mathbf{x}$ , the WP transform creates a tree with the nodes that correspond to the sub-bands of the appropriate scale.

As noted, the choice of a particular wavelet basis and of the sparsest subset of coefficients is very important with block signals. In particular, in the case of the WP representation, the best basis is chosen from the library of bases, and is used for decomposition, providing a complete set of features (coefficients).

In contrast, in the context of the signal source separation problem, it is useful to extract first an overcomplete set of features. Generally, in order to construct an overcomplete set of features, the approach to the BSS allows to combine multiple representations (for example, WP with various generating functions and various families of wavelets, along with DFT, DCT, etc.). After this overcomplete set is constructed, the 'best' subsets (with respect to some separation criteria) is chosen and form a new set used for separation. This set can be either complete or undercomplete. [29]

Alternatively, the following iterative approach can be used. First, one can apply the Wavelet Transform to original data, and apply a standard separation technique to these data in the transform domain. As such separation technique, one can use the NG approach, or simply apply some optimization algorithm to minimize the corresponding log-likelihood function. This provides us with an initial estimate of the unmixing matrix  $\mathbf{W}$  and with the estimated source signals. Then, at each iteration of the algorithm, we apply a multinode representation (for example, WP, or trigonometric library of functions) to the estimated sources, and calculate a measure of sparsity for each subset of coefficients. For example, one can consider the  $l_1$  norm of the coefficients  $\sum_{i,k} |c_{ik}|$ , its

modification  $\sum_{i,k} \log |c_{ik}|$ , or some other entropy-related measures. Finally, one can combine a new data set from the subsets with the highest sparsity (in particular, we can take for example, the best 10% of coefficients), and apply some separation algorithm to the new data. The iteration of the algorithm is completed. This process can be repeated till convergence is achieved.[29]

When signals have a complex nature, the approach proposed above may not be as robust as desired, and the error-related statistical quantities must be estimated. [29] use the following approach. Firstly, applying the Wavelet Transform to original data, and apply a standard separation technique (Natural Gradient, clustering, or optimization algorithm to minimize the log-likelihood function) to these data in the transform domain. Secondly, given the initial estimate of  $\mathbf{W}$  and the subsets of data (coefficients of mixtures) for each node, estimating the corresponding error variance, as described in[29]. Finally, one can choose a few best nodes (or, simply, the best one) with small estimated errors, combine their coefficients into one data set, and apply a separation algorithm to these data.

#### IV. CONCLUSIONS

In this article, we considered the problem of Multiresolution subband blind source separation (MRSBSS). We reviewed the feasibility of adaptively separating mixtures generated by the MRSBSS model and classified them into different categories, and identified new trends. After the perspective of MRSBSS is briefly presented, the general and detailed definition of the MSBSS model is given, the relationships among BSS, ICA and MRSBSS method is discuss simultaneously. Based on the minimization of the mutual information between outputs, we presented an adaptive algorithm for MRSBSS, which is called band-selective ICA. Then, taking the advantage of the properties of multiresolution transforms, such as wavelet packets, to decompose signals into sets of local features with various degrees of sparsity by the traditional NG method. Some simulations of these algorithms have been given to illustrate the performance of the proposed methods.

#### 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. Hyvarinen, J. Karhunen, E. Oja, Independent component analysis, John Wiley & Sons, New York, 2001.
- [2] A. Cichocki, S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, John Wiley&Sons, New York, 2002.
- [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] 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.
- [6] A. Hyvarinen, P.O. Hoyer, M. Inki, "Topographic Independent Component Analysis", *Neural Computation*, vol.13, no.7, pp.1527-1558, 2001.
- [7] F.R. Bach, M.I. Jordan, "Kernel Independent Component Analysis", *Journal of Machine Learning Research*, vol.3, pp.1-48, 2002.
- [8] K. Zhang, L.W. Chan, "An Adaptive Method for Subband Decomposition ICA", *Neural Computation*, vol.18, no.1, pp.191-223, 2006.
- [9] 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.
- [10] 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.
- [11] M.R. Aghabozorgi, A.M. Doost-Hoseini, "Blind separation of jointly stationary correlated sources", *Signal Processing*, vol.84, no.2, pp.317-325, 2004.
- [12] 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.
- [13] 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.
- [14] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol.36, no.3, pp.287-314, 1994.
- [15] M. Zibulevsky, B.A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary", *Neural Computation*, vol. 13(4), pp.863-882, 2001.
- [16] M. Zibulevsky, Y.Y. Zeevi, "Extraction of a source from multichannel data using sparse decomposition," *Neurocomputing*, vol.49, pp.163-173,2002.
- [17] F.M. Naini, G.H. Mohimani, M. Babaie-Zadeh, C. Jutten, "Estimating the mixing matrix in Sparse Component Analysis (SCA) based on partial k-dimensional subspace clustering", *Neurocomputing*, vol.71, no.10-12, pp.2330-2343, 2008.
- [18] P. Boffill and M. Zibulevsky, "Underdetermined blind source separation using sparse representation", *Signal Processing*, vol.81, no.11, pp.2353-2362, 2001.
- [19] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York, 1999.
- [20] M.V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A.K. Peters, 1994.
- [21] 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), vol.5, pp.541-544, 2004.
- [22] 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, vol.E86-A, no.1, pp.522-531, 2003.
- [23] 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.
- [24] 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.
- [25] O. Yilmaz, S. Rickard, "Blind separation of speech mixtures via timefrequency masking". *IEEE Transactions on Signal Processing*, vol.52, pp.1830-1847, 2004.
- [26] P.D. O'Grady, B.A. Pearlmutter, S.T. Rickard, "Survey of Sparse and Non-Sparse Methods in Source Separation ", *International Journal of Imaging System Technology*, vol.15, no.1, pp.18-33, 2005.
- [27] J.-F. Cardoso. Infomax and maximum likelihood for blind separation. *IEEE Signal Processing Letters*, vol.4, no.4, pp.112-114, 1997.
- [28] A. Prieto, C. G. Puntonet, and B. Prieto. A neural algorithm for blind separation of sources based on geometric properties. *Signal Processing*, vol.64, no.3, pp.315-331, 1998.
- [29] P. Kisilev, M. Zibulevsky, Yehoshua Y. Zeevi, "A Multiscale Framework For Blind Separation of Linearly Mixed Signals", *Journal of Machine Learning Research*, vol.4, pp.1339-1364, 2003.
- [30] R.W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Transactions on Image Processing*, vol.8, no.12, pp.1688-1701, 1999.
- [31] S. Amari, "Natural gradient works efficiently in learning", *Neural Computation*, vol.10, no.2, pp. 251-276, 1998.
- [32] J.F. Cardoso, B. Laheld, "Equivariant adaptive source separation", *IEEE Trans. Signal Processing*, vol.44, no.12, pp.3017-3030, 1996.
- [33] A. Cichocki, S. Amari, K. Siwek, T. Tanaka, ICALAB for signal processing; Toolbox for ICA, BSS and BSE. Available from <http://www.bsp.brain.riken.jp/ICALAB/ICALABSignalProc/>, 2004.
- [34] Fasong Wang, Hongwei Li and Rui Li, "Efficient Blind Source Separation Method for Harmonic Signals Retrieval", Accepted for publication.

**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.

**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.

**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, the 27th Research Institute, Zhengzhou, Henan, 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.