

# Adaptive Signal Separation with Maximum Likelihood

Yongjian Zhao\* and Bin Jiang\*

## Abstract

Maximum likelihood (ML) is the best estimator asymptotically as the number of training samples approaches infinity. This paper deduces an adaptive algorithm for blind signal processing problem based on gradient optimization criterion. A parametric density model is introduced through a parameterized generalized distribution family in ML framework. After specifying a limited number of parameters, the density of specific original signal can be approximated automatically by the constructed density function. Consequently, signal separation can be conducted without any prior information about the probability density of the desired original signal. Simulations on classical biomedical signals confirm the performance of the deduced technique.

## Keywords

Density, Estimator, Framework, Kurtosis, Likelihood, Separation

## 1. Introduction

In recent decades, blind signal processing (BSS) has been widely concerned for its potential applications in biomedical signal processing, automatic control, advanced statistics, and other academic and industrial fields [1-3]. Generally speaking, BSS technology is to obtain the relevant information we are interested in from the observed data of hybrid systems (such as wireless channel, communication system, radar system, mixing process, etc.) through some signal processing methods. The term “blind” denotes that all information about the mixing system is unknown in advance except for observed data. BSS is a typical neural network technique, whose goal is to estimate the original signals based on the observed signals without prior knowledge of the original signals and mixed parameters [2,3]. The BSS problem is mathematically underdetermined. Indeed, there are two uncertainties in the results of BSS separation. The first is the sequence uncertainty of the separation results and the second is the amplitude uncertainty of the separation results. The information to be transmitted is often contained in the signal waveform. Therefore, these two uncertainties do not affect the actual application of BSS technology. As a powerful statistical and computational technique, BSS has been proved to be reasonable and reliable in theory and will have great vitality [4-6].

Since Jutten and Herault [7] creatively released the original BSS approach in a typical feedback structure framework, a growing number of new theories and technologies have been put forward by follow-up researchers from various application fields such as digital images and economic indicators

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received June 26, 2018; first revision September 20, 2018; accepted September 25, 2018.

Corresponding Author: Bin Jiang (jian123cn@sdu.edu.cn)

\* School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, China ({zhaoyj, jian123cn}@sdu.edu.cn)

[2,8-11]. For instance, Virta and Nordhausen [8] proposed a blind signal separation method for multivariate time series. This method can be utilized to extract functional magnetic resonance imaging (fMRI) information from noisy high-dimensional observations. However, it works well only when the observed series is a linear transformation and is not correlated in time. If the period of desired source signal is close to another one, this method will lack robustness and reliability. Pehlevan et al. [9] formulated a non-negative signal blind separation scheme based on approximate similarity measurement. Indeed, such measurement has been utilized successfully in geophysical exploration. As a result, a corresponding explicit neural network approach is developed through the exploration of a typical similarity matching object. The local learning principle has been biologically verified and is widely used in many disciplines. The synaptic weights of the designed neural network are updated successively according to this learning principle. However, the proposed approach can only deal with a special case where expected original signals are confirmed to be non-negative in advance. This greatly limits its practical application. These typical BSS methods have opened a remarkable chapter in the history of signal processing. So far, BSS technology has been applied in many interdisciplinary fields due to its reliability and practicability [8,11].

The basic criteria for solving BSS problem include two kinds: second-order statistics and higher-order statistics [1,6,9]. Higher-order statistics contain a lot of information that second-order statistics do not own. Until now, higher-order statistics have shown stronger vitality than second-order statistics in signal detection, array processing, and object recognition [6,8,9]. Maximum likelihood (ML) is a powerful technique of higher-order statistics estimation [6]. The best way to separate signals in the BSS problem is to simulate the mixing process of original signals directly. However, it is difficult to know how the original signals are mixed in practice. Although the probability distribution of the observed mixture is unknown, it is a fixed value that exists objectively. A fundamental strategy for estimating conditional probability is to assume a certain probability distribution form and then optimize the parameters of the probability distribution according to the training samples [1,6]. The strategy of ML estimation is to estimate probability parameters based on data sampling. Such strategy makes it possible to separate the non-Gaussian component from its mixture successfully. The main appeal of ML estimation is its potential to develop into the best estimator asymptotically as the number of training samples approaches infinity [1,2]. As a powerful technique of statistical estimation, ML estimation is generally classified as the preferred estimator for BSS problem.

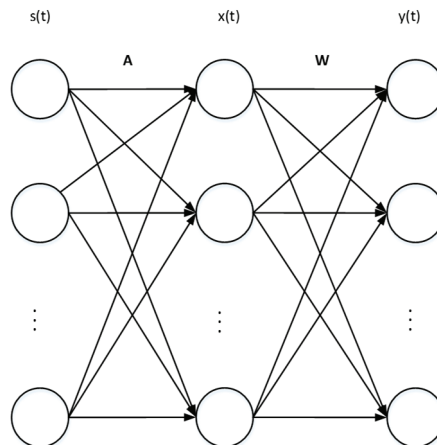
Under proper conditions, ML estimator has the property of consistency [1,2]. In other words, as the number of training samples tends towards infinitude, the ML estimation of a parameter converges to the true value of the parameter. Here, we regard ML as an attempt to match the probability density of the model to the real data that generate the probability density. Probability density function (PDF) is an expression representing the probability distribution of continuous random variables. An important question in BSS problem under ML framework is how to select a proper PDF to match the true data generating probability density distribution of the independent component (original signal). In other words, PDF plays a vital role for BSS problem in the ML framework. In essence, the estimation of probability densities is a typical nonparametric problem [2,10,11]. Although no direct access to the true data generating probability density distribution is available, efforts will be made to match this distribution to the greatest extent.

In this paper, a parametric density model is introduced correspondingly through a generalized exponential power family in the ML framework. In practice, source signals may have different kinds of

probability densities. It must be mentioned that the introduced density functions can be adaptive to various marginal probability densities after specifying a limited number of parameters. As a result, the probability densities of the original signals can be approximated automatically by the constructed density functions. Subsequently, a gradient learning algorithm for BSS problem is deduced in the ML framework by solving a constrained optimization problem. As a result, an adaptive method is deduced for BSS problem based on gradient optimization criterion. Compared with other methods in existence [7–9], the proposed method substantially differs in two aspects. Firstly, the signal separation can be carried out without any accurate prior information about the probability density of the desired original signal. Secondly, the proposed technique can separate different kinds of desired original signals from the observed mixture. The only assumption made for the signal separation process is that we should know in advance the kurtosis property of the desired original signal. Indeed, such signal kurtosis property can be readily known to the expert. Simulations on classical biomedical signals confirm the performance of the deduced technique.

## 2. BSS Problem Formulation

The fundamental BSS model can be represented in Fig. 1.



**Fig. 1.** The fundamental BSS model.

Suppose there are  $N$  independent original signals, expressed in vector form as  $s(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$ . Here superscript  $T$  represents the transpose of a vector and  $t = 0, 1, 2, \dots$ . We further suppose that there are  $M$  observed signals expressed as  $x(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$ . The  $M$  observed signals are obtained by the linear instantaneous mixing of  $N$  source signals. That is, at every moment there is the following relationship,

$$x_i(t) = \sum_{j=1}^N a_{ij} s_j(t) \quad i = 1, 2, \dots, M \quad (1)$$

Eq. (1) can be further rewritten as a vector matrix

$$x(t) = As(t), \quad (2)$$

where  $A$  is a mixing matrix composed of a series of mixing coefficients  $\{a_{ij}\}$ . The original signals and the mixing matrix are unknown, and only the mixed signals can be observed. Indeed, blindness is not complete since the original signals should be mutually independent in a statistical sense. In many cases, such as biomedical signal processing, only one specific original signal is of interest and the others can be ignored. Here, we focus on such practical application. Our main purpose is to obtain a separating vector and recover a particular original signal based on this vector from the observed mixtures. A fundamental strategy is to deduce an iterative process so as to obtain a separating vector  $W$ , which results in  $y = w^T x = w^T As$  being a scaled version of the interested original signal [12].

In the following, we only consider the case where the number of original signals is the same as the number of observed signals. Because if  $M > N$ , Principal Component Analysis (PCA) method can be used to remove redundant components [1,13,14]. It is also assumed that  $M = N$ , that is, the number of original signals is the same as the number of mixed signals. Because of the instantaneous mixing, the time variable  $t$  is omitted below in order to simplify the representation.

### 3. Proposed Algorithm

A crucial factor that constitutes the foundation of BSS is that original signals are statistical independent. In other words, the value of one variable cannot give any information on the value of another variable. Mathematically, statistical independence means that the joint probability density distribution of two random variables is equal to the product of their respective probability density distribution. In practical applications, signal independence may be measured with maximum entropy, mutual information, and ML [2,13,14]. ML estimator has properties of consistency and efficiency. ML becomes the best estimator asymptotically as the number of training samples increases and approaches infinity. While the number of samples is small enough to result in overfitting behavior, regularization strategies such as weight decay can be utilized to capture a biased version of ML that has less variance when training samples are limited [1,2,8]. In fact, ML is closely related to the approach of information flow maximization/minimization in neural networks. In recent decades, ML has become the preferred estimator to utilize for machine learning.

To perform ML estimation in practice, we should deduce a learning algorithm to carry out the numerical maximization/minimization of likelihood. The basic idea of BSS in the ML framework is to exploit the gradient of likelihood function with likelihood optimization criterion [3,15]. At the convergence point of the gradient optimization learning criterion, the gradient must point in the direction of  $W$ . In other words, the optimal gradient must be equal to the product of a constant scalar and the direction  $W$ . In such case, adding the gradient to  $W$  does not change its direction and the optimization algorithm converges here. Accordingly, the likelihood optimization criterion for separating a specific original signal in model (2) can be described by [11,16,17]:

$$\begin{cases} \max_w \psi(w) = E\{\log p(w^T x)\} \\ s.t. \quad \|w\|=1 \end{cases} . \quad (3)$$

Here  $p$  expresses the PDF of the original signal, which is unknown in practice and should be estimated in advance.

To solve the constrained optimization problem (3), we should calculate the optimal direction, which is mathematically the steepest direction. In practice, we can start with a particular vector  $W$ , calculate the direction in which  $x = As$  grows at the fastest speed based on available samples, and then turn  $W$  in that direction. This idea can be conducted in terms of the stochastic gradient optimization rule [1-3]. As a result, the following gradient learning algorithm can be deduced in the ML framework,

$$\begin{cases} w^+(i+1) = w(i) - \xi(i) E\{g(w(i)^T x)x\} / E\{g'(w(i)^T x)\} \\ w(i+1) = w^+(i+1) / \|w^+(i+1)\| \end{cases} \quad (4)$$

Here  $m$  expresses iteration index,  $\xi(i)$  is a step size related to  $i$ , and  $g(\cdot)$  describes the nonlinearity defined by  $g(\cdot) = (\log p(\cdot))' = p(\cdot)' / p(\cdot)$ . After the iterative process in (4) converges to a particular weight vector  $\tilde{w}$ , the specific interested original signal can be estimated with  $y = \tilde{w}^T x = \tilde{w}^T As$ . In the following, we call the algorithm in (4) MLBSS. Function  $p$ , which should match the true data generating probability densities of the independent components, plays a vital role when calculating the likelihood. In essence, the likelihood is a function of probability densities. These independent components are actually the source signals. How to choose the form of function  $p$  is an important and open research topic in the ML framework. If the true data generating probability densities lie within the constructed model family, the ML estimator will converge to the true results while the number of training samples approaches towards infinity [1,2]. The choice of function  $p$  is essentially a nonparametric problem, which contains an infinite number of parameters. A promising method is to select proper functions to approximate the densities of original signals [2,11].

**Theorem 1.** Suppose the probability density of original signal as  $\tilde{p}_i$ , and

$$g_i(s_i) = \frac{\tilde{p}_i'(s_i)}{\tilde{p}_i(s_i)} \quad (5)$$

The constraint of independent component estimation is set to be unit variance and uncorrelation. If supposed probability density  $\tilde{p}_i$  satisfies

$$E\{s_i g_i(s_i) - g'(s_i)\} > 0 \quad (6)$$

for all  $i$ , the ML estimator will be locally consistent.

Detail proof of Theorem 1 can be found in [1,11]. Theorem 1 shows that one can exploit a family of density functions that contain only two densities, thus making one of densities satisfy the conditions in (6). This means that one may make small mistakes in determining the density of independent component, since the estimated density can be guaranteed to lie within the same half of the probability density space. This also means that one can estimate the independent component with very simple density model.

Specifically, one can utilize a model which consists of only two densities.

In essence, ML can be considered an attempt to make the model probability density distribution match the empirical probability density distribution. Ideally, we would like to match the true data generating probability density distribution. Though lacking direct access, the true data generating probability density distribution must belong to the constructed model density family for otherwise no estimator can estimate the true data generating probability density distribution.

Here we consider the parameterized generalized distribution to match the probability density distribution of the interested original signal

$$p(y; \alpha) = \frac{\alpha}{2\beta\xi(1/\alpha)} e^{-\frac{|y|^\alpha}{\beta}} \quad (7)$$

where  $\xi(\cdot)$  is a Gamma function with form of  $\xi(m) = \int_0^\infty t^{m-1} e^{-t} dt$ , the Gaussian exponent  $\alpha$  is used to control the peakness of the density distribution. Since unit variance and uncorrelatedness mean  $E\{yy^T\} = wE\{xx^T\}w^T = I$ , we can further deduce  $\beta = \sqrt{\xi(1/\alpha)/\xi(3/\alpha)}$ .

In practical applications, the desired original signal may have deferent kinds of probability density distribution. In recent decade, there is a trend to exploit kurtosis as a gaussianity measure. For each iteration of learning algorithm (4), we can calculate the value of kurtosis. After convergence of this algorithm,  $w$  will be closely approximated to the optimal solution  $w^*$ , so the calculated kurtosis is narrowly approximated to its real value. Accordingly, the expected original signal can be estimated with  $y = w^*x$ . In practice, one may adjust the Gaussian exponent  $\alpha$  to control the peakness of the probability density distribution, thus making families in (7) consist of only two kinds of densities, i.e., a single binary parameter. With  $\alpha$  set to be various values, deferent kinds of probability densities ranging from super-Gaussian to sub-Gaussian are available. In other words, we can utilize parameterization of probability density distribution in (7), consisting of the choice between two densities.

Specifically, the MLBSS algorithm can be expressed as

**Step 1** Randomly initialize the separating vector  $W$ .

**Step 2** Calculate the kurtosis value.

**Step 3** Specify the Gaussian exponent  $\alpha$ .

**Step 4** Update the parameterized generalized distributions with Eq. (7).

**Step 5** Update the separating vector  $W$  with Eq. (4).

**Step 6** If vector  $W$  does not converge, go back to Step 2.

**Step 7** Separate the desired original signal with  $y = w^*x$ .

## 4. Computer Simulations and Performance Analysis

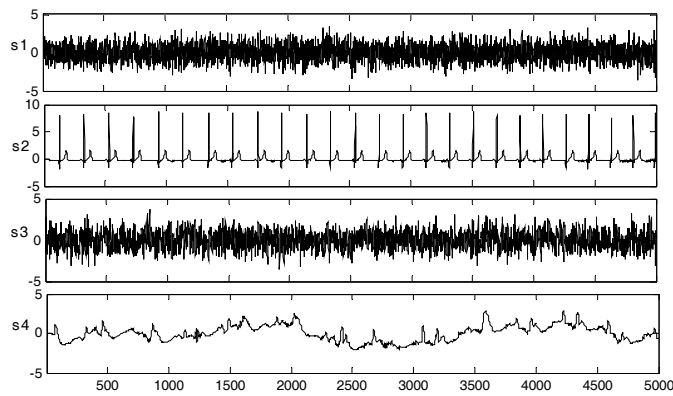
A main purpose of biomedical signal processing is to separate desired component from the observation of biosignal measurements that also contain uninterested signals. The ultimate objective is to extract clinically relevant information so as to improve the medical diagnosis. Fortunately, biomedical signals observed from a series of multiple measurements are statistically independent. Recently, there is a trend

to separate the desired biomedical signal from its observed measurements based on the BSS technique.

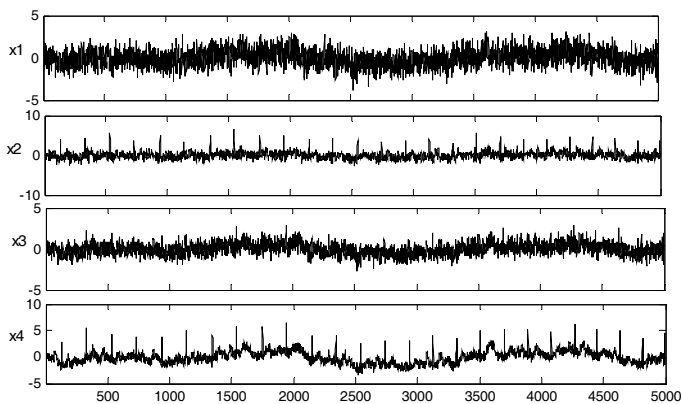
A nonsingular  $4 \times 4$  mixed matrix  $A$  was deduced randomly as

$$A = \begin{bmatrix} 0.7965 & 0.1021 & 0.1323 & 0.4024 \\ 0.6285 & 0.1982 & 0.2019 & 0.4088 \\ 0.1978 & 0.5987 & 0.5378 & 0.2968 \\ 0.3985 & 0.1098 & 0.4986 & 0.8956 \end{bmatrix}. \quad (9)$$

Four source signals as depicted in Fig. 2 were linearly mixed by matrix  $A$ . The corresponding mixing results were drawn in Fig. 3. As a graphic record of bioelectrical signals produced by the human body during cardiac circulation, electrocardiogram (ECG) may illustrate a lot about the medical condition of an individual. However, measured ECG is always contaminated by other signals or noise and is non-stationary in nature. It is imperative to analyze and interpret ECG signal with a powerful tool.



**Fig. 2.** Four source signals.

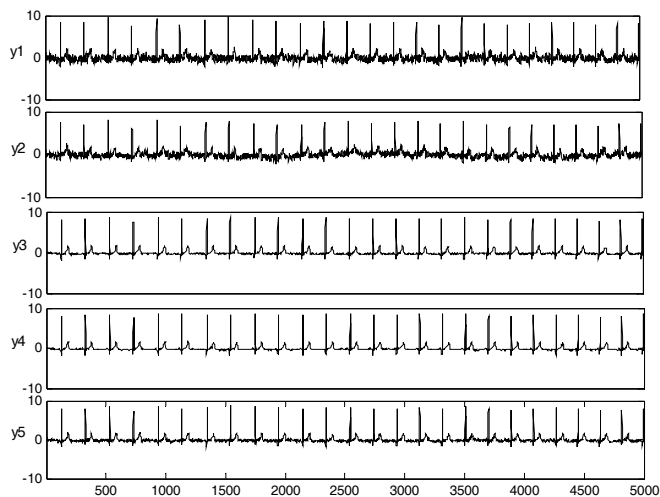


**Fig. 3.** Singles mixed by matrix  $A$ .

Our main goal was to separate a clear ECG from its signal mixtures exclusively. For comparison purpose, we ran three typical signal separation algorithms in sequence: heart sound segmentation BSS (HSSBSS) [5], constrained-linear-prediction BS (CLPBSS) [17], and our algorithm (MLBSS). ECG is a typical super-Gaussian signal. When MLBSS algorithm was conducted, Gaussian exponent  $\alpha$  in (7) was

set to 1, 1.6, and 3.0, respectively. All algorithms' parameters were adjusted in advance so that we could acquire the best average performance.

The signal separating results were drawn in Fig. 4. These results were in accordance with the sequence of the three separation algorithms. An important feature of HSSBSS algorithm is that it needs short data record because of the reduction of small sample separation error. One can find that signal  $y_1$ , separated by the HSSBSS algorithm, is always contaminated by other signals or noise. Functionally, the HSSBSS algorithm extracts source signals based on mutual information dependence and non-peak characteristics. Since the extracted original signal owns the biggest kurtosis value among all mixed source signals, it may have good performance for sound segmentation and extraction. Unfortunately, the desired ECG cannot satisfy this condition. The CLPBSS algorithm employs linear prediction to extract rhythm for signal separation, showing moderate separation performance. Linear prediction based method can extract source signals which have specific temporal structure. Fortunately, this type of information in the desired ECG is readily available. However, such method works well only when frequency and location of the interested original signal are available, which is not always realistic. In fact, the MLBSS algorithm has the best performance of the three algorithms. In particular, when  $\alpha$  is adjusted to 1.6, signal  $y_4$ , separated by the MLBSS algorithm, approximates to the original signal  $s_2$  to a great extent. Generally speaking, we note that  $\alpha=1.6$  is the best choice to separate the super-Gaussian signal from its mixture. In essence, data in ECG are often skewed due to larger signal amplitudes in activated regions, so exponential family can describe probability densities of the desired ECG effectively. In simulation, we found that one could adjust the Gaussian exponent  $\alpha$  to control the peakness of the probability distribution, thus making families in (7) consist of only two kinds of probability densities. In contrast, while the desired source signal was super-Gaussian, the Gaussian exponent value was set to 1.6 close to the best option; while the desired source signal was sub-Gaussian, the most appropriate variable approximates to 3.5. That is to say, the MLBSS algorithm can estimate different kinds of expected original signals. Most of all, as a signal separation algorithm in the ML framework, the signal processing can be carried out without any prior information about the probability density of the expected original signal. The MLBSS algorithm may separate a clear signal as long as its kurtosis property is known in advance.



**Fig. 4.** Separating results by various algorithms.



## 5. Discussion and Conclusions

ML estimation can transform the probability density estimation into a parameter estimation problem. As a fundamental method for statistical estimation, ML generally represents the preferred technique for BSS problem. In this paper, a learning algorithm, called MLBSS, has been proposed based on the stochastic gradient optimization rule for separating underlying component from source mixtures. A family of parameterized generalized distribution functions, which are adaptive to various marginal densities, have been deduced in this paper. One may set different exponential parameters, based on the kurtosis properties of desired source signal, to match different possible signal distributions. As a result, a gradient learning algorithm, which can separate different kinds of desired original signals, is deduced in ML framework. In fact, the MLBSS algorithm works in a semi-blind setting, since we should know the non-Gaussianity of the interesting original signal in advance.

In contrast to other ML based methods [1,2], the MLBSS algorithm has many advantages. Firstly, the existing ML based methods should know the probability density of the source signal in advance. The MLBSS algorithm can be implemented effectively as long as the kurtosis property of the expected original signal is known. Secondly, the existing ML based methods can only separate a few specific source signals. In contrast, the MLBSS algorithm may separate signals with super-Gaussian distribution or sub-Gaussian distribution, which is important in practice.

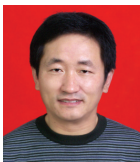
## Acknowledgement

This work is supported by Shandong Provincial Natural Science Foundation, China (No. ZR2017MA046).

## References

- [1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: The MIT Press, 2016.
- [2] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY: John Wiley & Sons, 2001.
- [3] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: from two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145-163, 2015.
- [4] R. Llinares, J. Igual, A. Salazar, and A. Camacho, "Semi-blind source extraction of atrial activity by combining statistical and spectral features," *Digital Signal Processing*, vol. 21, no. 2, pp. 391-403, 2011.
- [5] C. D. Papadaniil and L. J. Hadjileontiadis, "Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1138-1152, 2014.
- [6] H. Zhang, C. Wang, and X. Zhou, "An improved secure semi-fragile watermarking based on LBP and Arnold transform," *Journal of Information Processing Systems*, vol. 13, no. 5, pp. 1382-1396, 2017.
- [7] C. Jutten and J. Herault, "Blind separation of sources. Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1-10, 1991.
- [8] J. Virta and K. Nordhausen, "Blind source separation of tensor-valued time series," *Signal Processing*, vol. 141, pp. 204-216, 2017.

- [9] C. Pehlevan, S. Mohan, and D. B. Chklovskii, "Blind nonnegative source separation using biological neural networks," *Neural Computation*, vol. 29, no. 11, pp. 2925-2954, 2017.
- [10] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798-1828, 2013.
- [11] W. Y. Leong and D. P. Mandic, "Noisy component extraction (NoiCE)," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 3, pp. 664-671, 2010.
- [12] G. Chabriel, M. Kleinstueber, E. Moreau, H. Shen, P. Tichavsky, and A. Yeredor, "Joint matrices decompositions and blind source separation: a survey of methods, identification, and applications," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 34-43, 2014.
- [13] J. Nikunen and T. Virtanen, "Direction of arrival based spatial covariance model for blind sound source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 3, pp. 727-739, 2014.
- [14] Y. Zhao, B. Liu, and S. Wang, "A robust extraction algorithm for biomedical signals from noisy mixtures," *Frontiers of Computer Science in China*, vol. 5, no. 4, pp. 387-394, 2011.
- [15] M. Taseska and E. A. Habets, "Blind source separation of moving sources using sparsity-based source detection and tracking," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 657-670, 2018.
- [16] E. Santana, J. C. Principe, E. E. Santana, R. C. S. Freire, and A. K. Barros, "Extraction of signals with specific temporal structure using kernel methods," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5142-5150, 2010.
- [17] S. Ferdowsi, S. Sanei, and V. Abolghasemi, "A predictive modeling approach to analyze data in EEG-fMRI experiments," *International Journal of Neural Systems*, vol. 25, no. 1, article no. 1440008, 2015.



**Yongjian Zhao** <https://orcid.org/0000-0002-9203-8991>

He received the Ph.D. degree in biomedical engineering from Shandong University, Jinan, China, in 2009. He is currently an associate professor at Shandong University. His research interests include deep learning, signal processing, and pattern recognition.



**Bin Jiang** <https://orcid.org/0000-0003-2447-6946>

He received the Ph.D. degree from University of Chinese Academy of Sciences, Beijing, China, in 2012. He is an associate professor at Shandong University. His research interests include pattern recognition, data mining, and information security.