Minimum BER Criterion and Adaptive Moment Estimation Based Enhanced ICA for Wireless Communications

ZHONGQIANG LUO\textsuperscript{1,2}, (Member, IEEE), YAN CHEN\textsuperscript{1}, CHENGJIE LI\textsuperscript{3}, XINGZHONG XIONG\textsuperscript{1,} (Member, IEEE), AND LIDONG ZHU\textsuperscript{4}, (Member, IEEE)

\textsuperscript{1}Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Yibin 644000, China
\textsuperscript{2}Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things, Sichuan University of Science and Engineering, Yibin 644000, China
\textsuperscript{3}School of Computer Science and Technology, Southwest Minzu University, Chengdu 610041, China
\textsuperscript{4}National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Zhongqiang Luo (zhongqiangluo@gmail.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61801319, in part by the Sichuan Science and Technology Program under Grant 2020JDJQ0061, in part by the Education Agency Project of Sichuan Province under Grant 18ZB0419, in part by the Opening Project of Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things under Grant 2017WZZJ01, and in part by the Sichuan University of Science and Engineering Talent Introduction Project under Grant 2020RC33.

\textbf{ABSTRACT} This paper concentrates on investigating an enhanced independent component analysis (ICA) method for blind separation of signals corrupted by noise in wireless communications. Because of the traditional classical ICA methods that always have an inadequate capacity of anti-noise or insufficient separable ability in noise circumstance without satisfying practical application requirements. For this reason, two mechanisms are conducted to establish the modified cost function and fulfill the optimization assignment in the process of constituting an enhanced ICA algorithm. This proposed algorithm can benefit tremendously from the derived minimum bit error rate (BER) criterion and the novel adaptive moment estimation (Adam) optimization approach. In this work, firstly, the novel cost function is obtained according to minimum BER fused into maximum likelihood (ML) principle-based ICA cost function. Furthermore, by utilizing Adam processing, the task of blind separation of mixed signals is implemented via optimizing this modified cost function. Lastly, theoretical analysis and experiment results corroborate the better effectiveness and robustness of the proposed enhanced ICA algorithm compared with a series of popular representative ICA methods.

\textbf{INDEX TERMS} Independent component analysis, wireless communication, blind source separation, adaptive moment estimation, cost function, bit error rate, noise suppression.

\section{I. INTRODUCTION}

Blind source separation (BSS) has received attractive and remarkable attention from the academic and industrial community in recent decades. This data-driven signal processing method has played an increasingly significant role in numerous diverse disciplines, such as in wireless communications, biomedical sciences, speech signal separation, image processing, and so on [1]–[20]. Presently, the unsupervised learning based blind separation technology continues to be a research heated topic in the fields of signal processing and neural networks, especially in adaptive and intelligent information processing. Some researchers give top priority to solving some critical techniques while some others concentrate on the practical and valuable applications of BSS [1]–[4]. The BSS can extract or recover the unknown source signals from the observed mixed signals. Generally speaking, this blind character means the source signals and mixing system parameters are both unobserved in advance. Only a little of prior knowledge can be utilized for helping source recovery, usually statistical information. In real scientific applications, many of the observed mixed signals are modeled as a suite of sensors output, with receiving different linear combinations of the underlying source signals. Therefore, the interested
source signals are expected to be separated or extracted from the observed data directly with the aid of BSS exempting from extra parameter estimation, such as channel state information and synchronization parameters for wireless receiving processing [4], [7], [9], [10], [12]–[14].

It is noteworthy that an increasingly great deal of people has recognized the importance of taking advantage of the BSS theory for source recovery application owing to its variously convenient advantages. A typical example can be found in wireless communication, which benefits a lot from BSS by helping wireless communication system arrive at requirements of strong anti-interference and high spectral efficiency due to its blind and adaptive features for future green and intelligent communication implementation. In wireless communication systems, a number of receiving models can be constructed as a BSS framework or conceived as a BSS problem, such as DS-CDMA (direct sequence-code division multiple access) [7], [8], OFDM (orthogonal frequency division multiplexing) [9]–[11], MIMO (multiple input multiple output) [12]–[14] and wireless sensor network (WSN) [23]–[25], cognitive radio [26], and so on [4]. In a general way, those received models can be considered as mixtures of independent source and unknown channel condition. The expected signals can be separated or extracted from the received mixed signals by the independent component analysis (ICA) algorithm based BSS theory. By the BSS technique, several meaningful advantages will be achieved. For example, the repeatedly used pilot sequences can be eliminated for enhancing spectral efficiency. Besides, BSS can contribute significantly to resisting unpredictable interference and achieve performance enhancement of source recovery.

However, there is a striking problem concerning noisy model separation, which starts to be increasingly salient in the BSS field. Previous studies of noise suppression in the BSS model have not been dealt with entirely, and the noise term is always not considered in the BSS framework [1], [3], [4]. It must be acknowledged that noise is an inevitable adverse factor which extensively exists in the observed mixed signals, especially in wireless communication systems. Thus, the problem of effective blind separation of sources from the noisy mixture has given rise to a lot heated debated. The researchers have made tremendous efforts to improve the performance of blind separation of the noisy ICA model. Some investigators hold the view of adding the number of received sensors through noise estimation operation [1], [3], or constructing overdetermined model using principal component analysis (PCA) [7] for suppressing the influence of noise. But this processing will boost the complexity and the cost of the receiver. Some other researchers think that the preprocessing operation can be added before the blind separation work, for instance, wavelet denoising [16], empirical mode decomposition (EMD) [27]. Nevertheless, the separation task will become high computational complexity, and the separation performance has not yet attained the ideal effect as expected. Therefore, further investigations of solving this problem are strongly encouraged.

As a motivation, this article will give top priority to improving the essential of algorithm mechanism rather than just placing emphasis on previous thoughts about preprocessing and adding receiving antennas or sensors. Two proposed mechanisms will be harnessed for modifying the cost function and optimization operation for intensifying the effectiveness and robustness of anti-noise character. As is well-known, the ICA algorithms are composed of two steps, which directly affect the performance of the blind separation of source signals. Firstly, the cost function is constructed based on an independent principle. Secondly, the cost function is optimized for carrying out blind separation. Therefore, the incentive of remedy in those two steps is reasonable.

So far, there are three popular independence principles for establishing cost function, including maximum likelihood (ML), minimum mutual information (MMI) and non-Gaussian maximization. Some prestigious algorithms have been developed via those independence principles, such as FastICA, Infomax, and so on [1]–[4]. Those algorithms are always directly used to implement blind separation task for source recovery. The effectiveness and robustness of separation are weak in bad noise surrounding or even inseparable when the noise level is higher. This paper attempts to associate with the performance criterion of the communication system for constructing a hybrid cost function [12], [21]. In this article, the idea of a constrained BER criterion combined into ML or MMI principle is motivated to build the cost function of BSS.

Traditionally, when the cost function is established, then the stochastic gradient or natural gradient is used to optimize the built new cost function [1], [3]. However, these solutions will be often trapped in the trouble of slowly separating high-dimensional sources or bad scaled mixture condition. Moreover, the convergence speed of blind separation of the noisy mixture will become worse. Many methods have been developed to conquer this deficiency, for example, using a momentum term in the learning rule in [18], employing a self-adjusting variable step-size in [2], [17] or exploiting a scaled natural gradient algorithm in [3]. In this article, a novel algorithm called adaptive moment estimation (Adam) [19], [22] for gradient-based optimization will be exploited for improving its computational efficiency and convergence speed. This algorithm can reach a better separation performance thanks to using the mechanism of the adaptive moments estimation of the gradient.

In this work, an enhanced ICA (EICA) or E-ICA algorithm will be proposed combined with minimum BER criterion and Adam optimization approach. The derivation of the minimum BER criterion is illustrated in detail. Theoretical and simulation analysis will be implemented for verifying the effective and robust performance of the proposed EICA. The Amari Performance index (PI) and BER performance will be conducted to evaluate for validating the effectiveness of the proposed idea. Theoretical and experiments analysis demonstrate that the EICA algorithm exhibits better convergence.
and separation performance compared with that of the famous representative ICA methods.

The remainder of the article is constructed as follows. In Section II, the typical BSS system model is reviewed. The new cost function of ICA and the optimization mechanism for ICA algorithm are both described in Section III. Theoretical analysis and remarks are shown in Section IV. Simulation results and discussions are demonstrated in Section V. Conclusions are illustrated in Section VI.

II. SYSTEM MODEL

In this section, the basic BSS model is reviewed. As shown in Figure 1, the BSS model is closely related to the MIMO system. Of course, other communication systems, such as CDMA, OFDM, MIMO-OFDM, and Massive MIMO, can also be constructed as a BSS model after some other transformations. Considering the determined BSS model, that is to say, the number of transmitting sensors and receiving sensors is equal. Assume that the mutual independent source vector is denoted as \( \mathbf{S} = (s_1, s_2, \ldots, s_M)^T \). The mixing matrix \( \mathbf{A} \) represents the channel condition in a MIMO system. \( \mathbf{N} = (\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_M)^T \) is the noise vectors. The observed mixed signal is \( \mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M)^T \), in other words, the received signals in MIMO. The received mixed signals can be described as follows [1]–[4],

\[
\mathbf{X} = \mathbf{A} \mathbf{S} + \mathbf{N}. \tag{1}
\]

BSS aims to achieve blind separation or extraction source signals only from received signal mixtures. The source signals recovery can be obtained after the separating operation is executed,

\[
\mathbf{Y} = \mathbf{B} \mathbf{A} \mathbf{S} + \mathbf{B} \mathbf{N}. \tag{2}
\]

In an ideal case, \( \mathbf{C} = \mathbf{B} \mathbf{A} \) equals an identity matrix, i.e., the un-mixing matrix \( \mathbf{B} \) is an inverse form of the mixing matrix. However, in a real case, the matrix \( \mathbf{C} \) will be a generalized permutation matrix incurred from the inherent indeterminacy in BSS. It satisfies that this problem makes no difference to the separation assignment.

![FIGURE 1. The basic BSS model framework.](image)

III. METHOD

This section will develop an enhanced ICA via minimum BER criterion and Adam optimization for resisting noisy mixtures. Firstly, a minimum BER is derived for merging into ML or MML independent principle for constructing a new cost function. Secondly, the Adam approach is utilized to implement the optimization work of the cost function.

A. MODIFIED COST FUNCTION BASED ON MINIMUM BER CRITERION

ICA problems usually obtain its cost function through the maximum likelihood (ML) principle under the independence assumption. Suppose that sources \( \mathbf{S} \) are independent with marginal distribution \( f_{s_i}(s_i) \), thus the equation is satisfied as follows:

\[
f_{\mathbf{S}}(\mathbf{S}) = \prod_i f_{s_i}(s_i). \tag{3}
\]

In the traditional linear instantaneous ICA model, the noise term in equation (1) is neglected in the process of constructing cost function, i.e., \( \mathbf{X} = \mathbf{A} \mathbf{S} \). Thus the joint density of the observation vector and the joint density of the source vector can build the mathematical relationship as follows:

\[
f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{\det \mathbf{A}} f_{\mathbf{S}}(\mathbf{A}^{-1} \mathbf{X}) = \det \mathbf{A}^{-1} f_{\mathbf{S}}(\mathbf{A}^{-1} \mathbf{X}). \tag{4}
\]

To find a maximum likelihood estimation of \( \mathbf{A} \) (or \( \mathbf{B} \), where \( \mathbf{B} = \mathbf{A}^{-1} \)) is through maximizing the equation (4). Noting that \( \mathbf{Y} = \mathbf{B} \mathbf{X} \), the ML principle-based cost function can be derived from the logarithm likelihood of (4) as

\[
\log f_{\mathbf{X}}(\mathbf{X}) = - \log |\det \mathbf{A}| + \log f_{\mathbf{S}}(\mathbf{A}^{-1} \mathbf{X}). \tag{5}
\]

which can also be written as

\[
\log f_{\mathbf{X}}(\mathbf{X}) = \log |\det \mathbf{B}| + \log f_{\mathbf{Y}}(\mathbf{Y}). \tag{6}
\]

\( \mathbf{Y} \) is the estimation of with the true distribution \( f_{\mathbf{S}}(\mathbf{S}) \) replaced by a hypothesized distribution \( f_{\mathbf{Y}}(\mathbf{Y}) \). Due to sources have a statistically independent condition, so that the cost function is acquired as

\[
\mathcal{J}(\mathbf{B}) = - \log |\det \mathbf{B}| - \sum_{i=1}^{M} \log f_{y_i}(y_i). \tag{7}
\]

The cost function from the equation (7) can also be derived from the principle of mutual information minimum. The separation matrix is determined by

\[
\hat{\mathbf{B}} = \arg \min_{\mathbf{B}} \left\{ - \log |\det \mathbf{B}| - \sum_{i=1}^{M} \log f_{y_i}(y_i) \right\}. \tag{8}
\]

B. DERIVED MINIMUM BER CRITERION MERGING INTO TRADITIONAL ML BASED COST FUNCTION

This subsection will describe the minimum BER criterion firstly. Then ML principle-based cost function is built constrained by merging into this minimum BER criterion. In MIMO systems, the BSS problem is equivalent to blind equalization assignment. Considering the modulation digital signals in a MIMO system model, the equiprobable antipodal symbols are usually used for the transmitted symbols.
(for example, ±1, namely BPSK), and the uncorrelated with each other is assumed, i.e.,

$$E\left\{SS^T\right\} = I. \quad (9)$$

Due to the simplicity usage, we consider the antipodal assumption in derivation of the method. Note that we can extend it to the case of other constellations, for example, 4-QAM/QPSK. In addition, assume that the noise matrix $N$ conforms to a zero-mean, white and Gaussian, with covariance matrix as follows,

$$E\left\{NN^T\right\} = \sigma^2 I, \quad (10)$$

where $S$ denotes transmitted source signals, then $Y$, as given by the equation (2), represents the received signal matrix. In the next step, the matrix elements are quantized by a threshold detector to generate $Y_q$ with elements ±1. The average BER is defined as the average of the probability of error of each element of the block, yielding,

$$P_e = \frac{1}{M} \sum_{m=1}^{M} P_{em}. \quad (11)$$

In which $P_{em}$ denotes the BER of the $m$th source signal symbol. Consider that the signal power of each source symbol is assumed to be unity, then the covariance matrix of the received noise is $\sigma^2 BB^T$, which is obtained as follows,

$$E\left\{(S + BN)(S + BN)^T\right\} = E\left\{SS^T\right\} + BE\left\{NN^T\right\} B^T = I + \sigma^2 BB^T. \quad (12)$$

By further processing steps, the probability of the source symbol of $Y_q$ being in error is represented as,

$$P_{em} = \frac{1}{2} \text{erfc} \left( \frac{1}{\sqrt{2\sigma^2[BB^T]_{mm}}} \right). \quad (13)$$

In which $\text{erfc} (\cdot)$ $\triangleq 2/\sqrt{\pi} \int_{z}^{\infty} e^{-t^2} dt$, and $[BB^T]_{mm}$ represents the $(m,m)$th element of the matrix $BB^T$. $\sigma^2[BB^T]_{mm}$ is the received noise variance of the source symbol in the transmitted signal vector. Combined (13) with (11), yielding,

$$P_e = \frac{1}{2M} \sum_{m=1}^{M} \text{erfc} \left( \frac{1}{\sqrt{2\sigma^2[BB^T]_{mm}}} \right). \quad (14)$$

Assume that $\phi (z) = \text{erfc} \left( 1/\sqrt{2\sigma^2z} \right)$ for $z > 0$, we can achieve that

$$\frac{d^2 \phi}{dz^2} = \frac{1}{\sqrt{\pi}} \left( 2\sigma^2 \right)^{(1/2)} \exp \left( -\frac{1}{2\sigma^2z} \right) \left( -\frac{3}{2} + \frac{1}{2\sigma^2z} \right) z^{-3/2}. \quad (15)$$

Therefore, if $z < 1/3\sigma^2$, then $d^2 \phi/dz^2 > 0$. Applying this condition to the equation (14) $\phi \left( [BB^T]_{mm} \right)$ is a convex function when the noise power $\sigma^2$ is less than $1/3[BB^T]_{mm}$, when the condition is guaranteed for all $m$ (namely, the sufficiently large SNR (Signal Noise Ratio, SNR) exists at the receiver). In this case, the convex is also satisfied with the average block BER $P_e$. Due to that $P_e$ is convex at the condition of the moderate-to-high SNRs, we can use Jensen’s inequality to arrive at the following lower bound on $P_e$,

$$P_e = \frac{1}{2M} \sum_{m=1}^{M} \text{erfc} \left( \frac{1}{\sqrt{2\sigma^2[BB^T]_{mm}}} \right) \geq \frac{1}{2} \text{erfc} \left( \frac{1}{\sqrt{2\sigma^2} M \sum_{m=1}^{M} [BB^T]_{mm}} \right) = \frac{1}{2} \text{erfc} \left( \frac{1}{\sqrt{2\sigma^2} tr (BB^T)} \right) = P_e, LB. \quad (16)$$

The equality in (16) holds when all of $[BB^T]_{mm}$ are equal, $\forall m \in [1,M]$. The inequality (16) is valid, only when $P_e$ is convex [21], namely,

$$[BB^T]_{mm} < \frac{1}{3\sigma^2}, \quad \forall m \in [1,M]. \quad (17)$$

In (16), the $P_e, LB$ describes a lower bound on BER $P_e$. Since $\text{erfc} (\cdot)$ satisfies the monotonically decreasing character. Therefore, only minimize $tr (BB^T)$ can complete this purpose to minimize $P_e, LB$ in (16). Put it another way, the minimum BER criterion will be changed a description form as follows,

$$\min_{B} tr (BB^T), \quad s.t. \quad [BB^T]_{mm} < \frac{1}{3\sigma^2}. \quad (18)$$

Associated with (8), the modified cost function linked to the minimum BER can be constructed a constrained problem as,

$$\arg \min_{B} \left\{ -\sum_{i=1}^{M} \log (f_y (y_i)) - \log |\det B| \right\} \min_{B} tr (BB^T), \quad s.t. \quad [BB^T]_{mm} < \frac{1}{3\sigma^2}. \quad (19)$$

To simplify the constrained optimization problem in (19), the modified cost function via a minimum BER criterion in the moderate-to-high SNRs will be described as an unconstrained problem due to convex character, i.e.,

$$\mathcal{L} (B) = \left\{ -\sum_{i=1}^{M} \log (f_y (y_i)) - \log |\det B| \right\} + \lambda tr (BB^T), \quad \lambda > 0 \quad \lambda \in \mathbb{R} \quad (20)$$

where $\lambda$ is a regulation parameter, which conforms $|\lambda| < 1$. When the cost function is founded, then an appropriate method is adopted to optimize the cost function. In other words, the blind separation problem has been converted into a cost function an optimization problem. Traditionally, the optimization solutions are often carried out by the stochastic gradient or natural gradient.
C. MODIFIED COST FUNCTION THROUGH ADAM OPTIMIZATION FOR FULFILLING BLIND SEPARATION TASK

In this subsection, a novel gradient optimization approach is utilized to optimize the constructed cost function (20) [19], [22]. The traditional stochastic and natural gradient is also illustrated for introducing this new Adam optimization gradient method. Before using the natural gradient for $\mathbf{B}$, we should compute the gradient form to acquire the stochastic learning mechanism. The gradient of the cost function (20) related to the separating matrix $\mathbf{B}$ is derived as

$$
\frac{\partial \mathcal{L} (\mathbf{B})}{\partial \mathbf{B}} = -\frac{1}{M} \sum_{i=1}^{M} \frac{\partial \log \left( \tilde{f}_i (y_i) \right)}{\partial \mathbf{B}} - \frac{\partial \left[ \log (\det \mathbf{B}) \right]}{\partial \mathbf{B}} + \lambda \frac{\partial \text{tr}(\mathbf{B}^T \mathbf{B})}{\partial \mathbf{B}}.
$$

(21)

In equation (21), three terms of the gradient will be discussed respectively in the following. In ICA, to approximate a nonlinear function $g (.)$ for implementation. Regarding the first term in (21), the nonlinear function $g_i (.)$ has been incorporated into,

$$
\sum_{i=1}^{M} \frac{\partial \log \left( \tilde{f}_i (y_i) \right)}{\partial \mathbf{B}} = \sum_{i=1}^{M} \frac{1}{g_i (y_i)} \frac{\partial g_i (y_i)}{\partial y_i} \frac{\partial y_i}{\partial \mathbf{B}}
$$

$$
= \left( \begin{array}{cccc}
\frac{1}{g_1 (y_1)} \frac{\partial g_1 (y_1)}{\partial y_1} & \cdots & \frac{1}{g_1 (y_1)} \frac{\partial g_1 (y_1)}{\partial y_M} \\
\vdots & \ddots & \vdots \\
\frac{1}{g_M (y_M)} \frac{\partial g_M (y_M)}{\partial y_1} & \cdots & \frac{1}{g_M (y_M)} \frac{\partial g_M (y_M)}{\partial y_M}
\end{array} \right)
$$

$$
= \frac{1}{g(\mathbf{Y})} \mathbf{X}
$$

(22)

in which $g (\mathbf{Y})$ means $(g_1 (y_1), \ldots, g_M (y_M))$. Concerning the second term in (21), yielding,

$$
\frac{\partial \left[ \log (\det \mathbf{B}) \right]}{\partial \mathbf{B}} = \frac{1}{\det \mathbf{B}} \frac{\partial \det \mathbf{B}}{\partial \mathbf{B}}
$$

$$
= \frac{1}{\det \mathbf{B}} (\text{adj} (\mathbf{B}))^T
$$

$$
= \left( \mathbf{B}^{-1} \right)^T.
$$

(23)

In terms of the third term, we have,

$$
\frac{\partial \text{tr}(\mathbf{B}^T \mathbf{B})}{\partial \mathbf{B}} = 2 \mathbf{B}.
$$

(24)

Finally, the natural gradient update is generated as,

$$
\Delta \mathbf{B} = -\frac{\partial \mathcal{L} (\mathbf{B})}{\partial \mathbf{B}} \mathbf{B}^T
$$

$$
= \left( \mathbf{B}^T \right)^{-1} \mathbf{B}^T + \frac{1}{g(\mathbf{Y})} \frac{\partial g (\mathbf{Y})}{\partial \mathbf{Y}} \mathbf{X}^T \mathbf{B}^T \mathbf{B} - 2 \lambda \mathbf{B}^T \mathbf{B}
$$

$$
= \mathbf{B} + \varphi(\mathbf{Y}) \mathbf{Y}^T \mathbf{B} - 2 \lambda \mathbf{B}^T \mathbf{B},
$$

(25)

where

$$
\varphi (\mathbf{Y}) = \frac{1}{g (\mathbf{Y})} \frac{\partial g (\mathbf{Y})}{\partial \mathbf{Y}}.
$$

(26)

The ICA algorithm for blind separation assignment perform an iterative update of the matrix $\mathbf{B}$ as

$$
\mathbf{B}_t = \mathbf{B}_{t-1} + \mu \Delta \mathbf{B}.
$$

(27)

where the update step is given by (27) and $\mu$ denotes an update coefficient smaller than one, that affects the convergence speed of methods.

In the following, the Adam will be introduced firstly. Then the modified optimization approach via Adam algorithm will be proposed. Adam algorithm achieves adaptive estimates of lower-order moments, such as the first-order gradient-based optimization of stochastic objective functions [19], [22]. This method has several advantages, involving light memory requirements, straightforward implementation, high computational efficiency, and invariant to a diagonal rescaling of the gradient. Notably, this method is suitable for problems with very noisy and sparse gradients [19], [22]. In the following, we will talk about the fundamental mechanism of Adam algorithm.

Assume that a noisy cost function $f (\theta)$ is to be minimized related to the parameters $\theta$. The noisy gradient vector at a time $t$ of the cost function $f (\theta)$ related to the settings will be described as $\mathbf{g}_t = \nabla_{\theta} f_t (\theta)$. Adam algorithm executes the gradient descent (or ascent) optimization by assessing the square gradient $\mathbf{v}_t$ and the moving averages of the noisy gradient $\mathbf{m}_t$ [19], [22]. By using two scalar coefficients and which controls the exponential decay rates, these moment vectors are updated as follows:

$$
\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_t,
$$

(28)

$$
\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \mathbf{g}_t \odot \mathbf{g}_t,
$$

(29)

in which $\beta_1, \beta_2 \in (0, 1)$, and $\odot$ denotes the Hadamard multiplication. The initial values of $\mathbf{m}_0$ and $\mathbf{v}_0$ are zero vectors. These vectors indicate the mean and uncentered variance of the gradient vector $\mathbf{g}_t$. Since the estimates of $\mathbf{m}_t$ and $\mathbf{v}_t$ are biased towards zero, due to their initialization, a bias correction is computed on these moments [19]

$$
\hat{\mathbf{m}}_t = \mathbf{m}_t / (1 - \beta_1),
$$

(30)

$$
\hat{\mathbf{v}}_t = \mathbf{v}_t / (1 - \beta_2).
$$

(31)

This vector $\hat{\mathbf{v}}_t$ approximates as the diagonal of the Fisher information matrix. Hence Adam is strongly linked to the natural gradient algorithm. Lastly, the parameter vector $\theta_t$ is updated by the following rule at a time $t$,

$$
\theta_t = \theta_{t-1} - \eta \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t + \epsilon}}
$$

(32)

in which $\eta$ is a step size and $\epsilon$ is a small positive constant to overcome the zero division. In the case of gradient ascent, the minus sign in (32) is substituted with the plus sign.

Z. Luo et al.: Minimum BER Criterion and Adaptive Moment Estimation Based Enhanced ICA for Wireless Communications
In subsequence, the modified update rule is illustrated based on Adam optimization. It is noteworthy that Adam algorithm utilizes a vector of parameters, thus the vectorization of the gradient (25) is carried out, generating,

\[ \mathbf{b}_t = \text{vec} \left( \Delta \mathbf{B} \right) \in \mathbb{R}^{M^2 \times 1}, \]  
(33)

in which vec(·) is the vectorization operator, that obtains a vector by stacking the columns of the matrix below one another. The gradient vector is assessed on several blocks \( N_p \) extracted from the signals. At this case, the mean and variance vectors are estimated from the knowledge of the gradient \( \mathbf{b}_t \) at a time \( t \) by using equations (30) and (31). Then, using (32), the gradient vector (33) is updated for minimization of joint entropy by

\[ \mathbf{b}_t = \mathbf{b}_{t-1} - \eta \frac{\hat{\mathbf{m}}_t}{\sqrt{\mathbf{v}_t + \epsilon}}. \]  
(34)

At last, the vector \( \mathbf{b}_t \) is transformed in matrix form, by

\[ \mathbf{B}_t = \text{mat} \left( \mathbf{b}_t \right) \in \mathbb{R}^{M \times M}, \]  
(35)

in which mat(\( \mathbf{b} \)) rebuilds the \( M \times M \) matrix by unstacking the columns from the vector \( \mathbf{b} \). The whole algorithm is repeated for a certain number of epochs \( N_{ep} \). The pseudo-code of the minimum an enhanced ICA based on BER criterion and Adam modified ML, called here EICA or E-ICA, is illustrated in Table 1.

**TABLE 1. Pseudo-code illustration of EICA.**

| Algorithm/Pseudo-code Illustration of Enhance ICA |
|-------------------------------------------------|
| Data: Mixture Signals \( \mathbf{X} (n) \), \( \eta, \beta_1, \beta_2, \epsilon, N_B, N_{ep}, \lambda \); |
| 1 Initialization: \( \mathbf{B}_0 = \mathbf{I} \), \( \mathbf{m}_0 = \mathbf{0} \), \( \mathbf{v}_0 = \mathbf{0} \), \( \mathbf{b}_0 = \mathbf{0} \); |
| 2 \( P = N_B N_{ep} \); |
| 3 for \( t = 1 : P \) do |
| 4 Extract the \( t \)th block \( \mathbf{X}_t \) from \( \mathbf{X} (n) \); |
| 5 \( \mathbf{Y}_t = \mathbf{B}_{t-1} \mathbf{X}_t \); |
| 6 \( \varphi \left( \mathbf{Y}_t \right) = \text{tanh} \left( \mathbf{Y}_t \right) \); |
| 7 Update gradient \( \Delta \mathbf{B} \) in (25); |
| 8 \( \mathbf{b}_t = \text{vec} \left( \Delta \mathbf{B} \left( \mathbf{Y}_t \right) \right) \); |
| 9 \( \mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{b}_t \); |
| 10 \( \mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \mathbf{b}_t \odot \mathbf{b}_t \); |
| 11 \( \hat{\mathbf{m}}_t = \frac{\mathbf{m}_t}{1 - \beta_1} \); |
| 12 \( \hat{\mathbf{v}}_t = \frac{\mathbf{v}_t}{1 - \beta_2} \); |
| 13 \( \mathbf{b}_t = \mathbf{b}_{t-1} - \eta \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t + \epsilon}} \); |
| 14 \( \mathbf{B}_t = \text{mat} \left( \mathbf{b}_t \right) \); |
| 15 end |
| Result: Separated signals \( \mathbf{Y} (n) = \mathbf{B}_P \mathbf{X} (n) \). |

**IV. COMPLEXITY AND CONVERGENCE ANALYSIS OF EICA ALGORITHM AND PERFORMANCE EVALUATION**

It is noted that the separating matrix in a noisy mixed model is rather difficult or inaccurate to estimate. A majority of research efforts consider the noiseless case or the noises have a negligible effect on the performance of the algorithms. Nevertheless, the noise is always an inevitable factor in observed mixtures in the practical application. It is essential to separate the received mixed signal effectively and robustly, which plays a significant role in enhancing system performance.

This paper concentrates on promoting the blind separation performance of source signals corrupted by additive Gaussian noise. For achieving the adequate performance of separation signals, two modified notions are conceived to refine the separating performance of the traditional ICA method.

For one thing, a minimum BER criterion is investigated to merge into the traditional ML-based ICA cost function. This modified mechanism considers the lousy effect of noise term for constructing the cost function. In comparison, the initial ML principle-based cost function ignores the influence of the noise term. Nevertheless, the additive Gaussian noise is universal in many circumstances, especially in communications. Considering the update rule, the equation (25) can be rewritten as follows,

\[ \Delta \mathbf{B} = \mathbf{B} + \varphi \left( \mathbf{Y} \right) \mathbf{Y}^T \mathbf{B} - 2 \lambda \mathbf{B}^T \mathbf{B} \]

\[ = \Delta \mathbf{B}_{NG} - 2 \lambda \mathbf{B}^T \mathbf{B}. \]  
(36)

From the point of computation complexity, the designed cost function adds the \(-2\lambda \mathbf{B}^T \mathbf{B}\) term, and the increased computational complexity is \(O \left( M^3 \right)\) in comparison with that of the original \( \Delta \mathbf{B}_{NG} \). In terms of the question of convergence, we can derive,

\[ \| \Delta \mathbf{B} \|_2 = \| \Delta \mathbf{B}_{NG} - 2 \lambda \mathbf{B}^T \mathbf{B} \|_2 \leq \| \Delta \mathbf{B}_{NG} \|_2 + 2 \lambda \| \mathbf{B} \|_2 \| \mathbf{B}^T \|_2 \| \mathbf{B} \|_2 \]

\[ = \| \Delta \mathbf{B}_{NG} \|_2 + 2 \lambda. \]  
(37)

In the previous equation (37), the separating matrix is assumed to satisfy orthogonal property for simplified analysis. This condition is easy to meet due to that the whitening operation is always adopted to benefit from dealing with subsequent separating action conveniently. After the whitening processing, the mixing matrix attains the orthogonal matrix, and the separating matrix is the same as that case. From the regarding \( \lambda \) (\(|\lambda| < 1\)), it is a regulation parameter, is always enough appropriate small to balance the effect of the minimum BER criterion fused in the cost function of (36) while does not have the essential impact on the convergence of update rule. It must be acknowledged that \( \| \Delta \mathbf{B}_{NG} \|_2 \leq \delta \) is always considered as a convergence condition for the natural gradient algorithm. Because of the previous equation (37), we can acquire that the \( \| \Delta \mathbf{B} \|_2 \leq \delta \) will also be valid so that the proposed algorithm is validated to satisfy the convergence condition.

For another, for achieving computationally efficient and simplified, the Adam optimization method is utilized for finding the separating matrix. Adam approach [19], [22] has powerful and excellent properties that will benefit a lot the optimization problem of the cost function. This method has been shown that has \( O \left( \sqrt{T} \right) \) computation complexity (\(T\) denotes the sample size), which gets close to the known best bound for its general convex online learning. In contrast with the computational complexity of natural gradient \( O \left( M^2 T \right)\),
that of added complexity from the modified cost function may be neglected. Because the length of sample size is far larger than the number of source, i.e., $T \gg M$.

In conclusion, we can predict that the conceived method can carry out superior to the traditional ICA (not considered noise term) in separation assignment in terms of the noisy bad effect. In the next section, we will conduct simulation experiments to validate the effectiveness and robustness of the proposed method.

V. EXPERIMENT ANALYSIS AND DISCUSSIONS

For demonstrating the effectiveness and robustness of the designed algorithm, simulation experiments are conducted to evaluate the performance of the enhanced ICA (EICA or E-ICA). The performance of the Infomax algorithm [17], the Momentum Infomax (M-Infomax) [18], and Adam Infomax (A-Infomax) [19] are also illustrated for highlighting a comparative performance. The results are verified through the Amari Performance Index (PI) defined as follows [3] and Bit Error Rate (BER) which is an important performance index in communication systems.

$$PI = \frac{1}{M(M-1)} \sum_{i=1}^{M} \left( \sum_{j=1}^{M} \frac{|c_{ij}|}{\max |c_{ik}|} - 1 \right)$$

$$+ \sum_{j=1}^{M} \left( \sum_{i=1}^{M} \frac{|c_{ij}|}{\max |c_{ij}|} - 1 \right)$$

which $c_{ij}$ is the element of the matrix $C = BA$. This lower PI means better blind performance.

In a first experiment, five mixtures obtained as a linear combination of the following bad-scale independent sources are implemented from [19],

$$s_1[n] = 10^{-6} \sin(210n) \sin(60n),$$
$$s_2[n] = 10^{-5} \text{tri}(60n),$$
$$s_3[n] = 10^{-4} \sin(450n) \sin(80n),$$
$$s_4[n] = 10^{-5} \cos(400n + 4 \cos(50n)),$$
$$s_5[n] = \xi[n],$$

in which $\text{tri} (\cdot)$ is a triangular waveform and $\xi [n]$ represents a uniform noise in the range $[-1, 1]$. The simulation parameters refer to that of the literature of [19] to implement performance comparison conveniently. Each signal is composed of a sample size of $L = 30, 000$. The mixing matrix is generated as a $5 \times 5$ Hilbert matrix. A block length of $B = 30$ samples (hence $N_B = \lfloor L/B \rfloor = 1, 000$) is used, while the other parameters are set as $\beta_1 = 0.9, \beta_2 = 0.999, N_{ep} = 100$, $\varepsilon = 10^{-8}, \eta = 0.001, \lambda = 10^{-4}, \text{SNR} = 15 \text{dB}$, and the step size of the standard Infomax and Momentum Infomax to $\mu = 0.0001$, and the Adam Infomax is set as from that of the proposed EICA. BER performance (38) is portrayed in Figure 2, which easily shows the effectiveness of the proposed algorithm mechanism. The proposed enhanced ICA has better convergence performance than that of Adam Infomax, Infomax ICA, and Momentum Infomax thanks to incorporating the constrained rule and Adam optimization.

In a second experiment, assume that the number of transmitting antenna and receiving antenna is 4 in a MIMO system. The source symbols are generated from BPSK to make a BER performance comparison. The mixing matrix is generated from Gaussian distribution such as to simulate the AWGN channel. We use a block length of $B = 30$ samples, and the other parameters are set as $\beta_1 = 0.9, \beta_2 = 0.999, N_{ep} = 100$, $\varepsilon = 10^{-8}, \eta = 0.001, \lambda = 10^{-4}, \text{SNR} = 15 \text{dB}$, and the step size of the standard Infomax and Momentum Infomax to $\mu = 0.0001$, and the Adam Infomax is set as from that of the proposed EICA. BER performance is demonstrated in Figure 3, which illustrates the robustness of the proposed algorithm mechanism. The proposed enhanced ICA has better BER performance (separation accuracy) than that of Adam Infomax, Infomax ICA, and Momentum Infomax owing to incorporating the BER criterion constrained and Adam mechanism for constructing a cost function.

To further describe the superior performance of the proposed algorithm, the performance comparisons of direct blind separation exhibition are designed as shown in Figure 4 at
SNR = 12dB, the sample size of 1,000 data is illustrated in the figures. From the separated curves shown in Figure 4, we can safely conclude that the merging of the minimum BER criterion and Adam mechanism in BSS enhances the separation performance. It is noteworthy that the order of separation signal has been changed. But it is a common phenomenon that mainly comes from the inherent separation indeterminate problem of blind source separation. However, it won’t affect the separation performance. This issue is not our focus in this paper. It always can be overcome by utilizing some communication signal properties or sending short pilot sequence.

For further highlighting the performance of the conceived algorithm, Figure 5 and Figure 6 give a deep performance comparison of a series of representative algorithms in contrast with the proposed method. The BER of separation signals in a statistical analysis model is utilized to evaluate the performance. In the figure 6, the SNR is set as 9dB, and the other parameter is the same as that of in second simulation experiment, and the number of running experiments is 200. From the results of Figure 5, we can safely reach that the proposed E-ICA has improved performance compared to the other algorithms in separation performance. Moreover, in Figure 6, the error bar figure is illustrated to exhibit the performance comparison of different popular representative algorithms.

In a nutshell, these conducted simulation experiments have validated the robustness and effectiveness of the developed algorithm mechanism. Thanks to the modified BER
criterion-based cost function and Adam optimization facilitate the better separation performance for resisting the influence of the noise.

VI. CONCLUSION
This paper has put forward an enhanced ICA to carry out noise reduction to strengthen the separation performance in wireless communications. The proposed EICA algorithm takes advantage of the superiorities of the hybrid objective function and Adam stochastic optimization approach for implementing the improved effectiveness and robustness of blind separation of source signals. This modified ICA algorithm can achieve better performance in speed of convergence and separation accuracy in moderate SNR condition. It is strongly recommended to investigate the better constrained cost function and low complexity as well as fast optimization methods for the BSS problem in low SNR circumstance. Also, it is an attractive and promising research plan for considering other useful communication performance criteria or properties for helping develop advanced BSS algorithms for practical wireless receiving applications. And it is also interesting to investigate the theoretical performance bound for wireless signal separation and universal separation criterion from wireless signal mixture perspective.

ACKNOWLEDGMENT
The authors wish to express their gratitude to anonymous referees, whose insightful comments greatly helped improving the quality of the manuscript.

REFERENCES
[1] R. F. Coelho, V. H. Nascimento, R. L. de Queiroz, J. M. T. Romano, and C. C. Cavalcante, Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing, Boca Raton, FL, USA: CRC Press, 2016, pp. 3–23.
[2] X. C. Yu, D. Hu, and J. D. Xu, Blind Source Separation Theory and Applications. Singapore: Wiley, 2014, pp. 19–166.
[3] P. Comon and C. Jutten, Handbook of Blind Source Separation. Oxford, U.K.: Elsevier, 2010.
[4] Z. Luo, C. Li, and L. Zhu, “A comprehensive survey on blind source separation for wireless adaptive processing: Principles, perspectives, challenges and new research directions,” IEEE Access, vol. 6, pp. 66685–66708, 2018.
[5] C. Li, L. Zhu, and Z. Luo, “Underdetermined blind source separation of adjacent satellite interference based on sparseness,” China Commun., vol. 14, no. 4, pp. 140–149, Apr. 2017.
[6] X. Chen, Z. J. Wang, and M. J. Mckeown, “Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods,” IEEE Signal Process. Mag., vol. 33, no. 3, pp. 86–107, May 2016.
[7] Z. Luo and L. Zhu, “A charrelation matrix-based blind adaptive detector for DS-CDMA systems,” Sensors, vol. 15, no. 8, pp. 20152–20168, Aug. 2015.
[8] Z. Albutaiean and F. Salem, “Robust blind multiset detection algorithm using fourth-order cumulant matrices,” Circuits, Syst., Signal Process., vol. 34, no. 8, pp. 2577–2595, Aug. 2015.
[9] Z. Luo, L. Zhu, and C. Li, “Employing ICA for inter-carrier interference cancellation and symbol recovery in OFDM systems,” in Proc. IEEE Global Commun. Conf., Dec. 2014, pp. 3501–3505.
[10] L. Zhongqiang, Z. Lidong, and L. Chengjie, “Independent component analysis based blind adaptive interference reduction and symbol recovery for OFDM systems,” China Commun., vol. 13, no. 2, pp. 41–54, Feb. 2016.

Z. Luo et al.: Minimum BER Criterion and Adaptive Moment Estimation Based Enhanced ICA for Wireless Communications
YAN CHEN received the B.S. degree in communication engineering from the Sichuan University of Science and Engineering (SUSE), Zigong, China, in 2018. She is currently pursuing the master’s degree with the Sichuan University of Science and Engineering, Yibin. Her research interests include blind source separation, signal processing for wireless communication systems, and intelligent signal processing.

CHENGJIE LI received the B.Sc. degree from Shandong Normal University, Qufu, China, in 2004, the M.Sc. degree in computer software and theory from Xihua University, Chengdu, China, in 2009, and the Ph.D. degree in communication and information systems from the University of Electronic Science and Technology of China (UESTC), Chengdu, in 2017. Since 2017, he has been with Southwest Minzu University, where he is currently a Lecturer. His research interests include blind source separation, data mining, and intelligent information processing.

XINGZHONG XIONG (Member, IEEE) received the B.S. degree in communication engineering from the Sichuan University of Science and Engineering, Zigong, China, in 1996, and the M.S. and Ph.D. degrees in communication and information systems from the University of Electronic Science and Technology of China (UESTC), in 2006 and 2009, respectively. In 2012, he completed a research assignment at the Postdoctoral Station of Electronic Science and Technology, UESTC. He is currently a Professor with the School of Automation and Electronic Information, Sichuan University of Science and Engineering, Yibin. His research interests include information fusion, wireless and mobile communications technologies, intelligent signal processing, the Internet of Things technologies, and very-large-scale integration (VLSI) designs.

LIDONG ZHU (Member, IEEE) received the B.S. and M.S. degrees from Sichuan University, Chengdu, China, in 1990 and 1999, respectively, and the Ph.D. degree in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, in 2003. From July 2009 to June 2010, he was a Visiting Scholar with the Department of Electronics Engineering, The University of Hong Kong. He is currently a Professor with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China (UESTC). His research interests include satellite communications, blind source separation, and communication countermeasures.