A New Method of Blind Source Separation Using Single-Channel ICA Based on Higher-Order Statistics

Guangkuo Lu, Manlin Xiao, Ping Wei, and Huaguo Zhang

1Department of Information Engineering, University of Electronic Science and Technology of China, Chengdu, China
2College of Urban Railway Transportation, Shanghai University of Engineering and Science, Shanghai, China

Correspondence should be addressed to Guangkuo Lu; guangkuolu@gmail.com

Received 1 April 2015; Accepted 22 July 2015

Methods of utilizing independent component analysis (ICA) give little guidance about practical considerations for separating single-channel real-world data, in which most of them are nonlinear, nonstationary, and even chaotic in many fields. To solve this problem, a three-step method is provided in this paper. In the first step, the measured signal which is assumed to be piecewise higher order stationary time series is introduced and divided into a series of higher order stationary segments by applying a modified segmentation algorithm. Then the state space is reconstructed and the single-channel signal is transformed into a pseudo multiple input multiple output (MIMO) mode using a method of nonlinear analysis based on the high order statistics (HOS). In the last step, ICA is performed on the pseudo MIMO data to decompose the single channel recording into its underlying independent components (ICs) and the interested ICs are then extracted. Finally, the effectiveness and excellence of the higher order single-channel ICA (SCICA) method are validated with measured data throughout experiments. Also, the proposed method in this paper is proved to be more robust under different SNR and/or embedding dimension via explicit formulae and simulations.

1. Introduction

As one of the most attractive solutions for the blind source separation (BSS) problem, independent component analysis (ICA) has a strong practical background and wide applications in multiway data analysis such as biomedicine [1], image processing [2], telecommunications [3], geophysical research field [4, 5], and physics of musical instruments [6, 7], because it is a combination of informationism, optimal theory, probability, matrix theory, and mathematical statistics. Indeed, several papers recently have been written in which standard linear ICA, for instantaneous mixtures, was successfully applied to many natural environments as explosion-quakes and tremor at Stromboli and Erebos volcanoes [8, 9], to study acoustical and mechanical vibrational field in organ pipe [10]. Particularly, in intuitive way a segmentation combined with a ICA approach was already proposed to get information on the very long-period waves from water-level oscillations [11, 12], in which the ICA, the intertime occurrence, and the reconstruction of asymptotic dynamics are adopted after a preanalysis in the frequency domain. Obviously, ICA appears more appropriate in the investigation of nonlinear systems than the analyses based on the Fourier transform, even though several tidal behaviours have been pointed out by frequency-domain methods.

Generally, the number of sensors must be no less than that of the sources to acquire information to support the BSS work. Often in real cases, however, one has just a single measure of a certain specific physical variable, from which information on the underlying source mechanism has to be derived. In this case, the topic faced by the researchers is very important and difficult, that is, the extraction of characteristics from single experimental series, because of the lack of prior information. The method called SCBSS is proposed to exact the independent feature by using only one transducer. The methods employ ICA [13] to find the interested and independent feature from the decomposed signals based on oversampling [14], principal component analysis (PCA) [15], short time Fourier transform (STFT) [16], wavelet transform (WT) [17], empirical mode decomposition (EMD) [18, 19],
and so forth. The SCBSS methods based on oversampling and PCA can be used to separate the single-channel signal, however, only when the source signals in the linear mixture are stationary and independent. The methods based on STFT and WT can solve the nonstationary problem whereas they do not work to separate a nonlinear time series, which is generated from most of the natural [20, 21] and artificial systems in some fields like wireless communication and radar and sonar engineering. The SCBSS method based on EMD can be directly used to separate these single-channel signals, which are nonlinear, nonstationary, and even chaotic. But this method will break down if any source signal is not an Intrinsic Mode Function (IMF). Particularly, if the mixture contains spike pulses or the source signals have different time of arrival (TOA), which introduce some spurious extrema into the IMF confusion, and then a number of phantom sources can appear in the decomposed signals. To solve these problems, the way of thinking needs to be changed for developing new methods.

In a dynamical embedding framework, the measured data can be assumed to be generated by the nonlinear interaction of just a few degrees of freedom, with additive noise, and suggests the existence of an unobservable deterministic generator of the observed data. Obviously, in this case the reconstructed phase space (RPS) can be used to uncover as much information as possible about the underlying generators based only on the measured data [22], and the ICA algorithm could be then performed on the embedding matrix to extract its underlying ICs in the SCICA method proposed by James and Lowe [23, 24]. However, the SCICA method can successfully separate the measured signal only when the time series is nonlinear and stationary. Also, the key of the SCICA algorithm is to change the one-dimension time series to equivalent multidimensions through the RPS method. In order to achieve a better result, a larger embedding dimension should be taken, which could greatly increase computational complexity. To overcome this shortcoming, Ma [25] has developed a novel method, in which the stationary segments are firstly gotten by the Bernaola-Galavan (BG) segmentation algorithm [26], then the embedding dimension is reduced by singular spectrum analysis (SSA) [27], and the ICs are finally generated by ICA. The computational complexity of Ma’s method is successfully decreased from that of SCICA while achieving better performance. Unfortunately, although the selection of a suitable window length L in Ma’s method, which is crucial for the resolution of the SSA method to be computed by means of multiple autocorrelation, some subjective factors are introduced in the computational process. Essentially, SSA is a linear method based on the covariance matrix which reflects the linear relationship of the source signals and cannot reflect the intrinsic nonlinear relationship of them, although SSA-based method has been successfully applied in the field of signal processing for nonlinear dynamical systems. Particularly, the eigenvalues of the covariance matrix cannot be used to select a series of features to reconstruct the original time series if the signal to noise ratio (SNR) is too low or the embedding dimension is not correctly selected. Moreover, a multistage SSA algorithm [25] has been proposed to exact the feature signal under strong noise levels, which greatly increases the computational complexity.

In order to find a solution to the aforementioned problems, a modified method based on HOS is developed in this paper. Section 2 introduces signal model and the problem that needs to be solved in this paper. Section 3 contains the HOS-based SCICA method and simulations are carried out to verify the effectiveness of the method in Section 4. Finally, the conclusion of the paper is given in Section 5.

Notations. Hereinafter, bold uppercase letters denote matrices; bold lowercase letters stand for column vectors and lowercase letters represent scalars. Superscripts (,)T, |·|, and ‖·‖F denote transpose, absolute value, and Frobenius-2 norm, respectively. \( E[·] \) is the expectation operator, \( v_i \) is the \( i \)-th entry of \( v \). ∗ denotes convolution. \( \mathbb{R} \) denotes real number domain. \( \sigma \) and \( \tilde{\sigma} \) denote the true value and the estimate of variable \( \sigma \), respectively.

2. Problem Statement

2.1. Data Model. Generally, the observed single-channel signal \( x(t) \) could be modeled as a single-channel instantaneous linear mixture (SCILM) of \( n \) unknown independent signal \( s_k(t), k = 1, \ldots, n \):

\[
x(t) = a_1 s_1(t) + a_2 s_2(t) + \cdots + a_n s_n(t) + g(t),
\]

where \( a_k \) is the weight of source signal the \( k \)-th and \( g(t) \) is the zero-mean additive white Gaussian noise of unknown covariance. Obviously, ICA does not work when only one sensor could be employed. For the scenario, the key is to change the single time series into multidimensional time series before ICA will be used to separate the preprocessed signals. Based on different additional assumptions, the single-channel data can be reconstructed into different pseudo-MIMO models by different decomposition methods, such as PCA, STFT, WT, SSA, and EMD.

2.2. Single-Channel ICA. When the actual data is treated as a nonlinear time series with additive noise which is generated by the nonlinear interaction of just a few degrees of freedom, we can use the SCICA algorithm to solve the SCBSS problem. RPS is the first and foremost step, when the dynamic system theory is utilized to analyze a nonlinear time series. In [22], Takens shows that the map \( \Phi_{g,x} : M \to \mathbb{R}^{d+1} \) defined by \( \Phi_{g,x}(y) = (x(y), x(\varphi_1(y)), \ldots, x(\varphi_d(y))) \) is embedding, where the \( d \)-dimensional state space \( M, R \) is a twice continuously differentiable diffeomorphism that describes the dynamics of the system and \( \varphi \) is a twice continuously differentiable function representing the observation of a single state variable. Generally, the embedding dimension \( m \) must be large enough to capture the necessary information. Then for a nonlinear time series \( x(t) \in \mathbb{R} \), the state of the unobservable system at time \( t \), \( X(t) \) is given by

\[
X(t) = \{ x(t - \tau), x(t - 2\tau), \ldots, x(t - (m - 1) \tau) \},
\]

where \( \tau = p \cdot \Delta t \) is the lag, \( m \) is the embedding dimension, and \( \Delta t \) is the sampling interval.
Any approach to state space reconstruction uses the information in delay coordinates as a starting point. Obviously, Takens’ theorem allows us to reconstruct the unknown dynamical system that generates the measured time series by reconstructing a new state space based on the successive observations of the time series. It is indicated that the RPS of the nonlinear time series is the essential projection of the strange attractor on the axis of the space spanned by delay vectors. Therefore, each time series constructed by each delay vector can be regarded as a mixture of source signals. As shown in [23], the method based on RPS could be used to change the single-channel data into multidimensions time series. Then ICA can be used to span the embedding matrix with any ICs and to exact the feature.

2.3. Problem Statement. SCICA could separate a single-channel time series successfully if and only if this method satisfies the following conditions [23]:

(1) The measured signal is stationary.
(2) The phase state can be reconstructed perfectly.
(3) Each time series constructed by RPS could be considered as a single-channel instantaneous linear mixture (SCILM) of source signals.
(4) All the independent random processes must be bounded with disjoint spectral support.

Unfortunately, SCICA algorithms cannot be used directly for sources separation or extraction while the signal is nonstationary. Therefore, a nontrivial structure with nonstationarity of the actual signal with variable statistical property such as the mean and the variance is expected. The problem addressed in this paper is to segment a nonstationary time series, which consist of many segments with different statistical property, in such a way as to maximize the differences in the statistical property between adjacent segments. The BG algorithm in [26] is applied to divide the nonstationary data in [25]. However, an important assumption in BG algorithm that the variances of adjacent two segments are constant and the nonstationarity is only reflected by the difference of means of these two segments is not always true in a general sense. Therefore, the higher-order moments will be used for the nonstationary detection in this paper.

Takens’ theorem [22] shows that the unknown dynamical system can be reconstructed by recreating a new state space only when the Euclidean embedding dimension $m$ must satisfy that $m \geq 2d + 1$ ($d$ is the attractor dynamics). As shown in the proof of [22], the embedding dimension $m$ and the time lag $\tau$ could be selected arbitrarily, resulting in arbitrarily precise states, which is as good as any others. However, an important assumption of ‘Takens’ theorem is that the recording data without noise must be infinite, which may not always be true in the actual case. The actual data is always finite and is added with the strong broadband noise, which can obscure states and deteriorate the good properties of RPS. Simulation results [28] show that RPS does not work while the embedding dimension is less than the requirement. Accordingly, the calculated complexity increases with the increase of the embedding dimension. Although several methods, such as false nearest neighbor [29], singular value decomposition (SVD) [30], autocorrelation [31], and mutual information [32], can be used to determine the embedding dimension and the time lag of the reconstruction, these methods are mainly based on the experiments. Therefore, the selection of the reconstruction properties is essential to solve the problem of this paper.

Assuming instantaneous linear mixing of the sources at the sensors, ICA performs a blind separation of statistical independent sources with techniques involving higher-order statistics. However, RPS, which reconstructs the nonlinear time series in the state phase based on the delay coordinates, is essentially a nonlinear transform and cannot change single-channel data into multiple instantaneous linear mixture. Therefore, SSA [27] is used to transform the decomposed signals based on the delay coordinates into the one based on the original coordinates. In [25], SSA based on an eigenvalue decomposition (EVD) of the so-called lagged-covariance matrix for determining the optical dimension of reconstruction is applied to decompose the short and noise time series into a Pseudo-MIMO model before BSS is used. However, SSA, which essentially is a linear method, cannot reflect the structure of nonlinear dependence. Furthermore, SSA is not robust to reconstructive lag, embedding dimension, and the effect of the additive noise. Therefore, based on the above analyses, HOS-based methods may be employed because of the robustness of the higher-order-cumulants to Gaussian noise and the nonlinearity property.

3. Source Separation Using HOS-Based Single-Channel ICA

In this section, the actual data is assumed to be a stochastic process $x(t) = s(t) + g(t)$, where $s(t)$ is deterministic, $g(t) = h(t) * e(t)$, and $e(t)$ is an independent and identically distributed (IID) process. As shown in [33, 34] any process that satisfies the following assumptions can be referred to as a $k$th-order quasistationary process:

(1) $|s(t)| < \infty \forall t$,
(2) $E|e^{2N}(t)| < \infty$,
(3) $h(t)$ is linear, time-invariant, and stable; that is, $\sum_{\tau=-\infty}^{\infty} |h(\tau)| < \infty$,
(4) $\lim_{N \to \infty}(1/N) \sum_{t=1}^{N} E[x(t)x(t + \tau_1) \cdots x(t + \tau_{k-1})]$ exists and finite, $\forall \tau_1, \ldots, \tau_{k-1}$,

where $\tau_i$ ($i = 1, \ldots, k - 1$) is the time lag. Then, the fourth-order cumulants for a zero-mean, quasistationary (up to the fourth-order) signal $x(t)$ is defined as

$$C_{x_k}(\tau_1, \tau_2, \tau_3) = \text{cum}[x_{\tau_1} x_{\tau_2} x_{\tau_3}],$$

where $\text{cum}[s_1, s_2, s_3, s_4] = E[s_1 s_2 s_3 s_4] - E[s_1 s_2]E[s_3 s_4] - E[s_1 s_3]E[s_2 s_4] - E[s_1 s_4]E[s_2 s_3]$.

Furthermore, in this paper the actual data is considered as nonstationary signal, which is composed of many zero-mean, quasistationary (up to the fourth-order) segments
with different higher-order statistical properties. Then the different statistical properties will be selected to segment the time series into several subsets by means of the BG algorithm [26], which is based on heuristic segmentation with different scales and is more effective in detecting the abrupt changes of nonlinear time series.

Considering a zero-mean, quasistationary (up to the fourth-order) subset of the actual data, which is generated by a nonlinear dynamical system, the information about the underlying generators is uncovered by employing RPS-based method. Using the mean time between peaks (MTBP) as the time window, the reconstructed parameters $m$ and $\tau$ are estimated simultaneously and the nonlinear time series is reconstructed by Takens' embedding theory [22]. Then, the decomposition and reconstruction based on HOS are applied to reduce the dimension, weaken the noise, eliminate the nonlinear factors, and transform the phase space based on delay coordinates into the multiple instantaneous linear mixture. Finally, ICA is used to separate the decomposed sequences and extract the information from short and noisy time series. The modified SCICA strategy is illustrated in Figure 1, which will be further interpreted in the following sections in detail.

3.1. HOS-Based Segmentation of Nonstationary Time Series.

Since SCICA does not work to the nonstationary signals which are the property indeed of the measured single-channel actual signals, a modified BG algorithm is necessary for developing HOS-based methods. The BG algorithm is designed to characterize the stationary durations of human heart beat time series in [26]. The calculated complexity of this algorithm is reduced by iteratively segmenting the time series into only two segments. A decision to cut the time series is made by evaluating a modified Student’s $t$-test for the data in the two segments. However, the important assumption of this algorithm that the variances of both segments are kept constant and the nonstationarity is only reflected by the difference of means of the two segments may not always be true in a general sense as shown in Figure 2. Also, the BG algorithm does not work if the different scales are selected as shown in Table 1. In this paper, we suppose that the subsets composing the nonstationary signal are the stationary, real-valued, random processes with different means, and their moments up to order $n$ exist. Based on this assumption, a modified BG algorithm based on HOS is proposed in this section.
The original algorithm is modified as follows: After selecting a larger value as the minimum segment length, a sliding pointer is moved from left to right along the signal. At each position of the pointer, the time slots are computed as

$$T(i) = \left| \mu_i(i) - \mu_r(i) \right|,$$  \hspace{1cm} (4)

$\mu_i$ and $\mu_r$ are the mean of the subset of the signal to the left of the pointer and to the right, respectively. And, the statistical significance of $T_{\text{max}}$ is computed as

$$P(T_{\text{max}}) = \text{prob}(T \leq T_{\text{max}}),$$  \hspace{1cm} (5)

which could not be expressed in a closed form. Then, a suitable approximation by means of Monte Carlo simulations in [26] is given:

$$I_x(z, \omega) = \frac{1}{\beta(z, \omega)} \int_0^x t^{x-1} (t-1)^{\omega-1},$$

$$\beta(z, \omega) = \int_0^1 t^{z-1} (t-1)^{\omega-1} dt = \frac{\Gamma(z) \Gamma(\omega)}{\Gamma(z+\omega)},$$

$$\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt.$$  \hspace{1cm} (6)

This significance needs to be checked whether it is larger than a selected threshold $P_0$, which usually is taken to be 0.5–0.95. If so, the signal is cut at this point into two subsequences; otherwise the signal remains undivided. This process is continued to be conducted for each of these two subsequences recursively until the signal is composed of many minimum subsets. In other words, the process stops when none of the significance of the possible cutting points is larger than $P_0$. It needs to be noted that the process also stops if the length of subset is shorter than $l_0$ and we say that the signal has been segmented at the significance level $l_0$. Then the time series is expressed as

$$y(t) = x(t) - \mu(t),$$  \hspace{1cm} (7)

where $\mu(t)$ is the mean of the subset $i$. After selecting a smaller value as the new minimum length $l_0$, a sliding pointer is moved again from left to right along each subset segmented on different means. At each position of the pointer the time spots need to be computed as

$$T_i = [\sigma_i(i) - \sigma_r(i), s_i(i) - s_r(i), k_i(i) - k_r(i)]^T,$$  \hspace{1cm} (8)

where $\sigma(i), s(i),$ and $k(i)$ are variances, skewnesses, and kurtosises of the subsets of signal on both sides of the pointer. Since the statistical significance of $P(T_{\text{max}})$ cannot be reflected by the change of $T(i)$, BG algorithm can be used to segment the signal into subsets with different features corresponding to the 1st–4th order moments of time series.

### 3.2. HOS-Based Dynamic Systems Analysis of Nonlinear Time Series

The zero-mean, quasistationary (up to the fourth-order) segments prepared for SCICA could be exacted by a modified BG algorithm, as discussed above. Unfortunately, SCICA cannot be directly used to recover all sources from the recorded mixtures, in particular scenarios that the useful sources cannot be perfectly reconstructed into a phase state and the mixing segment cannot be considered as a SCILM. Applied to the actual data, therefore the reconstruction parameters can be selected by a new method to ensure validity of the RPS, and the reconstructed phase state in the delay coordinate system can be transformed into a multipath instantaneous linear mixture, which can be solved by ICA.

#### 3.2.1. RPS

The performance of RPS depends on two parameters, namely, the selection of the embedding dimension $m$ and the lag $\tau$. The correct estimations of these two parameters make the phase state be reconstructed perfectly. Recently some studies [35, 36] prove the relationship of these two parameters, which are expressed as

$$\tau_w = (m - 1) \tau,$$  \hspace{1cm} (9)

where $\tau_w$ is the time window.

Paper [35] shows that the nonlinear dynamics is successfully reconstructed in the phase space when the time window $\tau_w$ is kept constant and the other parameters are variable. Now we emphasize on how to use a simple way to estimate $\tau_w$, which equals the mean period $\tau_p$ in some applications. Strictly speaking, chaotic system is nonperiodic. For the low dimensional chaotic system with pseudo period, the mean orbital period is approximately equal to mean oscillation period of the chaotic attractors but never meets the orbit period of phase space. Meanwhile, the literatures [36] point out that the mean orbital period and the MTBP of chaotic time series are equal in general. The fast Fourier transform (FFT) can be employed to get the main frequency $f_s$ and the MTBP of the original time series is calculated to be the time window $\tau_w$, that is, $\tau_w = \tau_p = \text{MTBP} = 1/f_s$. Then the signal processed by RPS should provide a sufficient frequency coverage $f_w$ that can include all the frequency components related to the feature. Therefore, the following condition must be satisfied:

$$f_w = \frac{1}{\tau_w} \leq f_{L},$$  \hspace{1cm} (10)

where $f_{L}$ is the lowest frequency of interest. The sampling frequency meets

$$f_s = \frac{1}{\Delta t}.$$  \hspace{1cm} (11)

The sufficient frequency coverage for RPS is selected based on

$$f_w = \frac{1}{\tau_w} = \frac{1}{(m - 1) \tau} = \frac{1}{(m - 1) \Delta t} f_s \leq f_{L}.$$  \hspace{1cm} (12)
Then,
\[ m \geq \frac{f_c}{p f_L} + 1. \]  
(13)

Therefore, the lag can be set as small as 1 while the embedding dimension is large enough based on Takens’ theory, which is based on
\[ \tau_w = (m - 1) \tau = (m - 1) p \Delta t. \]  
(14)

If the embedding dimension \( m \) and the time lag \( \tau \) are obtained, the initial time series is changed into the trajectory matrix \( X \):
\[
X = \begin{bmatrix}
    x_t & x_{t+\tau} & \cdots & x_{t+(N-(m-1)\tau)} \\
    x_{t+\tau} & x_{t+2\tau} & \cdots & x_{t+(N-(m-2)\tau)} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{t+(m-1)\tau} & x_{t+m\tau} & \cdots & x_{t+N\tau}
\end{bmatrix}
\]  
(15)

Particularly, paper [35] shows that the value of the time lag could be reflected by the autocorrelation of signal. If the degree of autocorrelation of signals becomes lower, the smaller lag can be selected. Thus the time lag cannot be too large in the RPS for the signal with low SNR. Although the signal of RPS may be with low SNR because of the large time window, the modified method based on HOS in this paper could be applied to reduce the white noise.

3.2.2. HOS-Based Coordinate Transformations. For some purposes, such as reducing the dimension of the reconstruction, it may be desirable to make a further coordinate transformation \( \Psi \) to a new coordinate system [37]: \( Y = \Psi(X) \).

It is shown that performing SVD for the trajectory matrix \( X \) embedded by RPS is equivalent to apply EVD to the lagged-covariance matrix \( A_x \), which can be denoted as
\[
A_x = (X - \mu(X))(X - \mu(X))^T
\]
\[
= \begin{bmatrix}
    R_0 & R_\tau & \cdots & R_{(m-1)\tau} \\
    R_\tau & R_0 & \cdots & R_{(m-2)\tau} \\
    \vdots & \vdots & \ddots & \vdots \\
    R_{(m-1)\tau} & R_{(m-2)\tau} & \cdots & R_0
\end{bmatrix}
\]  
(16)

Then
\[
A_x(i, j) = E \left[ x(n)x(n+(i-j)\tau) \right] = R_{(i-j)\tau},
\]  
(17)

After the step of embedding, the EVD is applied to the lagged-covariance matrix \( A_x \), which is rewritten as
\[
A_x = VDU^T \left( VDU^T \right)^T = \begin{bmatrix}
    v_1 & v_2 & \cdots & v_m
\end{bmatrix}
\]
\[
\begin{bmatrix}
    \lambda_1 & 0 & \cdots & 0 \\
    0 & \lambda_2 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & \lambda_m
\end{bmatrix}
\begin{bmatrix}
    v_1^T \\
    v_2^T \\
    \vdots \\
    v_m^T
\end{bmatrix}.
\]  
(18)

In this step, we select \( d = \max(i : \lambda_i > 0) \) and the lagged-covariance matrix \( A_x \) can be denoted as
\[
A_x = \sum_{i=1}^{d} \psi_i \lambda_i \psi_i^T.
\]  
(19)

Then, a nonlinear transformation \( \Psi : \mathbb{R}^m \rightarrow \mathbb{R}^d \) is defined as
\[
Y = \Psi(X).
\]  
(20)

Although SVD can be used to construct a nonlinear transformation that (in a certain sense) provides the optimal coordinates in the RPS, unfortunately, singular spectrum obtained from some system data is not used for distinguishing the dynamic signal from the random noise. Furthermore, the autocorrelation-based methods cannot be robust enough for the embedding dimension and the lag with large dynamic range, which can reflect the trajectory matrix and the singular spectrum. Finally, the autocorrelation-based method, which is essentially a linear analysis technique, is difficult to reflect the intrinsic nonlinear correlation on space-time of nonlinear time series. Also, the observed data is composed of many zero-mean, quasistationary (up to the fourth-order) segments. Then to solve the above-mentioned problem, the algorithms based on HOS are considered to be used, which include the following advantages: (1) suppressing additive white and colored Gaussian noise of unknown power spectrum; (2) extracting information due to deviations from Gaussianity; and (3) detecting and characterizing nonlinear properties in signals as well as identifying nonlinear systems [38, 39].

Thus by applying the high-order cumulants theory, a modified method is provided. A fourth-order cumulant is defined as
\[
C_{4x} = \text{cum} \begin{bmatrix}
    x_t & x_t & x_t^T \\
    x_{t+\tau} & x_{t+\tau} & x_{t+\tau}^T \\
    \vdots & \vdots & \vdots \\
    x_{t+(d-1)\tau} & x_{t+(d-1)\tau} & x_{t+(d-1)\tau}^T
\end{bmatrix}
\]  
(21)
Then, the elements from the 4th-order cumulants function are selected as the elements of \( A_x(i, j) \). However the 4th-order cumulants function has three variables and cannot be directly used to SVD. To obtain a binary function, the 4th-order cumulants function is taken to many different slices, the \( A_x(i, j) \) can be reexpressed as

\[
A_x^{(0)}(i, j) = C_{4x}(i, j, j) = m_{4x}(i, j, j) - m_{2x}(i) \cdot m_{2x}(j) - 2m_{2x}(j) \cdot m_{2x}(j - i) + m_{3x}(j, j) + 2m_{3x}(j, j),
\]

where \( m_{nx} \) is the \( n \)-th order moment function for a zero-mean, quasistationary (up to \( n \)-th order) signal \( x(t) \). When the observational noise complicates the reconstruction problem for the actual data. When a \( d \)-dimensional state is projected onto a \( m \)-dimensional measurement with \( m < d \), the information is thrown away. However, if the uncertainty of the reconstructed state is much higher than that of the individual measurements, the noise is amplified. In that case, the change of the properties of the reconstruction \( m \) and \( \tau \) will result in the change of the trajectory matrix \( X \) and even cause the change of the diagonal matrix \( D \) and the new coordinate system \( Y \) as shown in Figure 3. This phenomenon is referred to as noise amplification, which can be defined as

\[
\sigma^2(T) = \lim_{\varepsilon \to 0} \frac{\text{Var} (Y(T) | x)}{\text{Var} (x)} = \lim_{\varepsilon \to 0} \frac{1}{\text{Var} (x)} \int \left( \int p(x | y) dx \right) \left( \int y^2 p(x | y) dx \right),
\]

where \( p(x | y) = p(x - \Psi(y(T))) \) and \( \varepsilon \) is the given noise level.

Obviously, the noise amplification depends on the measurement function, the method of reconstruction, and the dynamical system as proved in paper [40]. It means that the method of SCBSS for nonlinear time series is not robust for the different SNR and the different factors such as \( m \) or/and \( \tau \). Thus the noise amplification has to be defined to compare different methods of state space reconstruction. This gives guidance for optimizing the parameters of a particular method, or for comparing two different methods.

Then we show that selecting the optimal slice is equivalent to minimizing the noise amplification \( \sigma(T) \), which can be estimated from the time series \( x(t) \). For noise reduction, the true value is defined as \( s(t) = a_1s_1(t) + a_2s_2(t) + \cdots + a_ms_m(t) \). Since \( x(t) = s(t) + g(t) \), the true noise amplification can be defined as

\[
\tilde{\sigma}^2(T) = \sigma^2(T) - 1.
\]

Obviously the true noise amplification \( \tilde{\sigma}^2 \) can be used as evaluative criteria for the cumulants slices. Then, the smaller the value of \( \tilde{\sigma}^2 \), the better the robustness of the selected cumulants slice for SNR, \( m \), and \( \tau \).

As shown in the following experiments, there are several general motivations behind the use of the HOS-based method in signal processing.

(1) This method is used to distinguish the feature signal from the Gaussian noise. Clearly, consider signal \( x(t) = a_1s_1(t) + a_2s_2(t) + \cdots + a_ms_m(t) + g(t) \), where \( s(t) \) is the zero-mean, deterministic signal, and \( g(t) \) is the Gaussian noise. When \( x(t) \) is quasistationary up to the fourth order, then the fourth-order mixed cumulant is given by

\[
C_{4x}(r_1, r_2, r_3) = C_{4x}(r_1, r_2, r_3) + C_{4g}(r_1, r_2, r_3)
\]

Because of more effectiveness for suppressing additive Gaussian noise and reflecting the intrinsic nonlinear correlation in nonlinear time series, the 4th-order cumulants will be used for conducting the singular spectrum analysis of nonlinear
time series in this paper. Figure 4 illustrates the results of singular spectrum analysis obtained by a cross-correction method and a technique based on cross fourth-order cumulants. The signal of interest is assumed to be nonlinear and non-Gaussian whereas the additive noise is Gaussian. From Figure 4, it is apparent that the HOS-based coordinate transformation does suppress the effect of Gaussian noise and thus provides better singular spectrum, especially in low SNR.

(2) As shown in Figure 5, the method based on high-order moments is a more robust technique for autocorrelation-based method under different embedding dimensions, which is the greatest superiority for the nonlinear time series analysis.

(3) The nth-order cumulants function of a non-Gaussian stationary random signal \( x(t) \) can be written as (for only)

\[
C_{m,n}(\tau_1, \tau_2, \ldots, \tau_{n-1}) = m_{m,n}(\tau_1, \tau_2, \ldots, \tau_{n-1}) - m_{m,n}^{eq}(\tau_1, \tau_2, \ldots, \tau_{n-1}),
\]

where \( m_{m,n}(\tau_1, \tau_2, \ldots, \tau_{n-1}) \) is the \( n \)-th-order moment function of \( x(t) \) and \( m_{m,n}^{eq}(\tau_1, \tau_2, \ldots, \tau_{n-1}) \) is the \( n \)-th-order moment function of an equivalent Gaussian signal that has the same mean value and autocorrelation sequence as \( x(t) \). The third motivation is based on the measured signals in the real-world, most of which are nonlinear and non-Gaussian and have nonzero higher-order moments. For the nonlinear system analysis, the HOS-based coordinate transformation can reflect the third-order correlation of the nonlinear time series while the autocorrelation-based methods only give expression to the linear correlation. Thus more information about the nonlinear feature of the nonlinear time series is exacted as shown in Figure 6.

The HOS-based method is used to process the trajectory matrix \( X \) corresponding to the measured single-channel signal \( x(t) \) by means of RPS. A typical process of HOS-based optimal coordinate transformation has the following stages.

(1) Calculate the fourth-order correlation function \( C_4(x_1, x_2, \ldots, x_L) \) and construct the Toeplitz matrix \( A_x(i, j) \) with the optimal slice \( C_4(x_1, x_2, \ldots, x_L) \).

(2) Apply SVD on the Toeplitz matrix \( A_x(i, j) \):

\[
[V, D] = A_x.
\]

(3) Project the trajectory matrix \( X \) on the subspaces \( V_j \):

\[
Y_j = V_j V_j^T X^T,
\]

where the subspace spanned by the \( j \) eigenvectors \( V_j \) is \( V_j V_j^T \). Then, the trajectory matrix \( X \) can be written as

\[
Y = Y_1 + Y_2 + \cdots + Y_L.
\]

(4) Reconstruct the original series \( X \) corresponding to the selected features based on the inverse Hankelization by antidiagonal averaging into a new timeseries of length \( N \), and
obtain the time series $X_j$ from averaging of the corresponding diagonals of the matrix $Y_j$:

$$x_j(k) = \begin{cases} \frac{\sum_{m=1}^{k} y_j(m,n)}{k}, & 1 \leq k \leq L - 1, \\ \frac{\sum_{m=L}^{N} y_j(m,n)}{L}, & L \leq k \leq N - L + 1, \\ \frac{\sum_{m=L-N}^{N-k} y_j(m,n)}{N - k + 1}, & N - L + 2 \leq k \leq N, \end{cases}$$

where $m = 1, 2, \ldots, k$, $m + n = k + 1$. Then we obtain the pseudo-MIMO model: $X = [x_1; x_2; \ldots; x_L]$, where $L$ is the number of feature components, which need to be further separated by means of the ICA method.

### 3.3. Independent Component Analysis

Then, ICA [13] is chosen to represent the data in the pseudo-MIMO model in the assumption that (a) the measured data provided by the HOS-based method can be considered as a linear instantaneous mixing, and (b) any source of interest is independent of the other signals. ICA, which is a blind separation of statistically independent sources, performs as a means of identifying underlying components of the actual signals data. This method yields more useful results than other methods such as PCA as shown in [41].

In essence, ICA must find a separating or demixing matrix $W$, which is sufficient to $Y = WX$, where the components in the separated signals $Y$ are as independent as possible. The general method based on the minimization of mutual information [42, 43] can be described as

$$w_i(k + 1) = \frac{w_i(k + 1)}{\|w_i\|_2},$$

where $z$ is the whitening data, $w_i$ is the row vector of demixing matrix $W$, $k$ is the iterative sequence number, $E$ indicates mean value, and $g$ is the score function, which can be obtained as gradient of entropy $\hat{H}(Y)$. $\hat{H}(Y)$ can be estimated by

$$\hat{H}(Y) = -\sum \hat{\pi}(i) \log \hat{\pi}(i) - \log \det h,$$

where $\hat{\pi}(i) = (1/k) \sum_{n=1}^{k} \delta[i - h^{-1}Y(n)]$, $k(u) = \begin{cases} \frac{3}{4} - \frac{u^2}{2}, & |u| \leq 1/2 \\ \frac{3/2 - |u|^2}{2}, & 1/2 \leq |u| \leq 3/2 \\ 0, & \text{otherwise}, \end{cases}$

and $h$ is a smoothing parameter matrix. Then the estimation of the probability density $p_Y$ is

$$p_Y(y) = \frac{1}{k} \sum_{n=1}^{k} \delta[h^{-1}(y - Y(n))],$$

and the mutual information can be estimated as

$$\tilde{I}(y_1, \ldots, y_k) = \sum_i \tilde{\pi}_Y(i) \log \frac{\tilde{\pi}_Y(i)}{\prod_{k=1}^{K} \tilde{\pi}_{Y_k}(i_k)},$$

![Figure 5: Singular values of the measured signal: (a) HOS-based method; (b) autocorrelation-based method.](image)
where $\hat{r}_{Y_k}(j) = \sum_{i:j=i} \hat{r}_{Y_k}(i_k)$. Therefore, the separated signals can be described as

$$Y = WZ.$$  (37)

where $Y$ is the separated independent data.

Performing ICA on the matrix processed by HOS-based method $X$ results in a set of ICs that forms the basis of the matrix. Particularly, since the data has been processed by HOS-based coordinate transformation, the ICs as many as measurement channels can be directly considered as all of the principal components in the data. Furthermore, the ICs of interest can be identified and exacted.

### 3.4. The Implementation of the Algorithm

Based on the previous sections, we can introduce a modified SC-ICA algorithm by the following implementation:

1. Detect and segment the measured signal into several high-order stationary subsets (with the modified BG segmentation algorithm) in time domain.
2. Determine the window $L$ (based on the time window method) and transform each subset into multidimensional time series $X$ (based on the RPS).
3. Perform the HOS-based coordinate transformation to change the trajectory matrix $X$ into the pseudo-MIMO model.
4. Separate the source signals with ICA.

### 4. Simulations

A single-channel observation of two sources is taken as an example. The sampling frequency $f_s$ is set to 1 GHz. The data consists of 30,000 samples and the signal is cut into a series of quasistationary (up to the fourth-order) subsets by means of the modified BG algorithm. A subset which consists of 4096 samples is taken as a time series under test. The power spectrum of the mixture of sources is shown in Figure 7. Note that just from the power spectrum someone may be fooled into believing that there are three distinct sources present. After calculating the minimum time window $t_w$ and determining the embedding dimension $m = 50$ with a lag $\tau = 1$, the time series can be reconstructed into the multidimensions time series $x$. Applying autocorrelation-based method to trajectory matrix $X$ results in the singular
Figure 8: Singular spectrum of time series by means of autocorrelation-based method.

Figure 9: Singular spectrum of time series by means of HOS-based method.

The spectrum as shown in Figure 8. Figure 8 also shows that the number of singular values associated with the signal subspace is three. Obviously, the autocorrelation-based method does not work. On the contrary it is clear that the HOS-based method successfully recognizes the number of the sources shown in Figure 9. Then, the number of path is selected as six and the power spectrum of the mixture of sources by means of RPS and HOS-based coordinate transformations is shown in Figure 10. Obviously, the data from six channels can be regarded as multipath instantaneous linear mixture model. Finally, the two sources are successfully separated via ICA as shown in Figure 11, and the other ICs generated by ICA are the useless ones as shown in Figure 12.

5. Conclusion

Overall, this paper presents an approach for exacting information from single real-world data. The idea is firstly to...
segment the measured signal and then to form a pseudo-MIMO system by means of decomposing the observed segment into several signals using a representation method based on higher-order statistics (HOS). Finally the fixed-point FastICA algorithm is applied to estimate the source signals (independent components). The simulations show that the method is successful in isolating components from the single-channel data. Also, the methods based on fourth-order cumulants are more robust than those based on autocorrelation as the properties of the reconstruction are changed. Compared with the autocorrelation method, HOS-based SCICA is better for low SNR. At this stage, HOS will not be more sensitive to the number of the used samples. This is a problem for all the HOS-based method when they are applied to the actual data, but it is not a problem for this method. Moreover, since the available techniques used in this paper can process the single-channel signal without depending on a priori, the method is very powerful method that can isolate feature components in the actual data.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61201282) and the Fundamental Research Funds for the Central Universities of China (no. ZYGX2013J016).

References


