Gradient method for blind chaotic signal separation based on proliferation exponent*

Lü Shan-Xiang(吕善翔), Wang Zhao-Shan(王兆山), Hu Zhi-Hui(胡志辉), and Feng Jiu-Chao(冯久超)†

School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, *China*

(Received 13 June 2013; revised manuscript received 24 July 2013; published 2 December 2013)

A new method to perform blind separation of chaotic signals is articulated in this paper, which takes advantage of the underlying features in the phase space for identifying various chaotic sources. Without incorporating any prior information about the source equations, the proposed algorithm can not only separate the mixed signals in just a few iterations, but also outperforms the fast independent component analysis (FastICA) method when noise contamination is considerable.

Keywords: blind separation, chaotic signals, phase space

PACS: 05.45.−a, 05.40.Ca, 05.45.Vx **DOI:** 10.1088/1674-1056/23/1/010506

1. Introduction

In past decades, the studies on nonlinear dynamical systems, particularly chaotic systems, have drawn much attention. Chaotic signals characterized by aperiodicity and sensitivity to initial conditions, by nature, feature random behaviors. The discovery of chaotic synchronization has drawn close attention to applications of chaos in communications and signal processing.[1–3] However, in time-varying, multi-path, and fading channels where multiuser systems may apply, the issues of channel interference and mixed multiuser carriers require researchers to tackle the blind source separation (BSS) problem.

Much work has been done to separate chaotic signals by means of coupled synchronization,[4,5] particle filter,[6] or Kalman filter.[7,8] However, potential applications of these methodologies are confined to some scenarios where the equations of the chaotic maps are known by transcendental knowledge. As for the rest, algorithms based on independent component analysis (ICA) have been articulated in Refs. [9]–[11]. Basically, what characterised ICA is that it grabbles components that are both statistically independent and non-Gaussian.[12] Among the ICA algorithms, the fixed-point method searching for fourth-order cumulants called fast independent component analysis (FastICA) provides a radically faster learning rate,[13] while other algorithms using maximum likelihood estimation[14] or minimization of mutual information[15] seem to be mundane and mediocre. In addition, the value of noise variance is often presumed when applying the ICA methods to noisy data, albeit the actual physical noise level acquired through measuring instruments is a pronumeral; and because of this, the hypothetical noise level is fussy.[16,17] These results motivate us to study a new objective function and corresponding gradient-type algorithm for chaotic signal separation.

This paper aims to address the blind separation problem from a new perspective without any prior knowledge about the chaotic signals. Starting from phase-space reconstruction, the proliferation exponent (PE) of a time series is defined to achieve a simple quantitative description for chaotic signals. By setting PE as a cost function, its gradient-type algorithm is derived to deal with the blind separation task. The remainder of this paper is organized as follows. The proliferation exponent is defined and its properties are given in Section 2. The detailed algorithm to solve the separation problem is presented in Section 3. A few examples are also given to illustrate the performance of our algorithm in Section 4. Finally, Section 5 ends with conclusions.

2. Proliferation exponent

2.1. Definition

Takens has proposed a delay-embedding method for phase-space reconstruction which constitutes the foundation of chaotic time series analysis.[18] Time-delay embedding is certainly not the only method for embedding scalar data into a multi-dimensional dynamical system. Nevertheless, time-delay embedding is perhaps the only systematic method which transforms scalar data to multi-dimensional phase space and, to a large extent, has been explored extensively in the literatures. Considering an observable $s(t)$, by using a collection of time lags to create a vector with d dimensions, we have

$$\boldsymbol{s}(t) = [s(t), s(t+\tau), s(t+2\tau), \ldots, s(t+(d-1)\tau)]^{\mathrm{T}}, \quad t = 1, 2, \ldots, \sigma+1, \tag{1}$$

*Project supported by the National Natural Science Foundation of China (Grant No. 60872123), the Joint Fund of the National Natural Science Foundation and the Natural Science Foundation of Guangdong Province, China (Grant No. U0835001), the Fundamental Research Funds for the Central Universities of China (Grant No. 2012ZM0025), the South China University of Technology, China, and the Fund for Higher-Level Talents in Guangdong Province, China (Grant No. N9101070).

†Corresponding author. E-mail: fengjc@scut.edu.cn

© 2014 Chinese Physical Society and IOP Publishing Ltd

where d is the embedding dimension, τ is the time lag, and $\sigma+1$ is the amount of vectors in the phase space.

In the context of blind signal processing, how shall we choose the time lag τ and the dimension of the space d to go into a time-lagged vector $s(t)$? Takens suggested that it does not matter what time lag one chooses, while a sufficient condition for d depends on the dimension of the attractor d_{A}. They argued that if d is an integer and larger than $2d_{\mathrm{A}}$, which can be fractional, then the attractor as seen in the space with lagged coordinates will be smoothly related to the attractor as viewed in the original physical coordinates. In practice, their theorem inspires us that, if the value of d we choose is large enough so that it approaches the length of a securable signal, then the physical properties of the attractor which we wish to extract from the measurements should be identical to those of the original.

With the help of phase-space reconstruction, a distance function of $\boldsymbol{s}(t+1)$ to $\boldsymbol{s}(1)$ is defined as

$$V_s(t) = ||\boldsymbol{s}(t+1) - \boldsymbol{s}(1)||^2, \tag{2}$$

where $||\cdot||^2$ denotes the square of norm.

As different chaotic signals correspond to various growth rates in the phase space, the definition of proliferation exponent of a scalar time series is hereby regulated as follows:

$$P_\sigma(\boldsymbol{s}) = \frac{\mathbb{D}(V_s(t))}{\mathbb{E}^2(V_s(t))}, \qquad t = 1,2,\ldots,\sigma, \quad \sigma \in \mathbb{Z}, \tag{3}$$

where $\mathbb{D}(\cdot)$ denotes variance, $\mathbb{E}(\cdot)$ stands for expectation, and σ is the control factor. Intrinsically, equation (3) intends to measure the growth rate of the compact manifold, in which embedded dimension $d \gg d_{\mathrm{A}}$ and time lag $\tau = 1$.

The value of control factor σ affects not only the precision of the growth rate of $V_s(t)$, but also impresses the convergence speed of its corresponding blind separation algorithm. Basically, if the measurement along a chaotic trajectory is clean, then we may specify the control factor to be $\{\sigma | 5 \leq \sigma \leq 10, \sigma \in \mathbb{Z}\}$ that measures the initial fluctuation region of chaotic maps or flows, hence reducing the computational burden for $P_\sigma(\boldsymbol{s})$. However, experimental measurement on a physical system is often polluted by noise, so achieving statistical analysis for the consecutive sample points forbids too small σ value in these scenarios. In these applications, the σ value for observables from noisy physical systems can be set coarsely as $\{\sigma | \sigma \geq 100, \sigma \in \mathbb{Z}\}$ because of their small sampled intervals. A larger σ value, by nature, results in a smoother objective surface of PE.

2.2. Property and comparison

Given the definition, PE features the following properties.

Property 1 Scale invariance, i.e., $P_\sigma(\boldsymbol{s}) = P_\sigma(k\boldsymbol{s}), \forall k \neq 0$.

Proof As $V_{k\boldsymbol{s}}(t) = k^2 V_s(t)$, substituting it into Eq. (3), we have

$$P_\sigma(k\boldsymbol{s}) = \frac{k^4 \mathbb{D}(V_s(t))}{k^4 \mathbb{E}^2(V_s(t))} = P_\sigma(\boldsymbol{s}).$$

Property 2 Upper bound, i.e., $P_\sigma(k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j) < \max(P_\sigma(\boldsymbol{s}_i), P_\sigma(\boldsymbol{s}_j))$, $\forall k_1, k_2 \neq 0$, where $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ denote irrelevant variables.

Proof Let $P_\sigma(\boldsymbol{s}_i) < P_\sigma(\boldsymbol{s}_j)$, then we need to prove

$$\frac{\mathbb{D}(V_{k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j}(t))}{\mathbb{E}^2(V_{k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j}(t))} < \frac{\mathbb{D}(V_{k_2\boldsymbol{s}_j}(t))}{\mathbb{E}^2(V_{k_2\boldsymbol{s}_j}(t))}. \tag{4}$$

Since $\mathbb{D}(\boldsymbol{s}(t+1) - \boldsymbol{s}(1)) = \sigma||\boldsymbol{s}(t+1) - \boldsymbol{s}(1)||^2$, by taking advantage of the irrelevant feature of $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, we can write $V_{k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j}(t)$ as

$$\begin{aligned} V_{k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j}(t) &= \frac{1}{\sigma}\mathbb{D}(k_1\boldsymbol{s}_i(t+1) - k_1\boldsymbol{s}_i(1) \\ &\quad + k_2\boldsymbol{s}_j(t+1) - k_2\boldsymbol{s}_j(1)) \\ &= V_{k_1\boldsymbol{s}_i}(t) + V_{k_2\boldsymbol{s}_j}(t). \end{aligned} \tag{5}$$

For the sake of brevity, note $V_{k_1\boldsymbol{s}_i}(t)$ and $V_{k_2\boldsymbol{s}_j}(t)$ as X and Y, respectively. From Eq. (5) we can derive

$$\begin{aligned} \frac{\mathbb{D}(V_{k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j}(t))}{\mathbb{E}^2(V_{k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j}(t))} &= \frac{\mathbb{D}(X) + \mathbb{D}(Y)}{\mathbb{E}^2(X) + \mathbb{E}^2(Y) + 2\mathbb{E}(X)\mathbb{E}(Y)} \\ &< \frac{\mathbb{D}(X) + \mathbb{D}(Y)}{\mathbb{E}^2(X) + \mathbb{E}^2(Y)}. \end{aligned} \tag{6}$$

Since

$$\frac{\mathbb{D}(X) + \mathbb{D}(Y)}{\mathbb{E}^2(X) + \mathbb{E}^2(Y)} < \frac{\mathbb{D}(Y)}{\mathbb{E}^2(Y)}, \tag{7}$$

the proposition of Eq. (4) is correct according to Eqs. (6) and (7).

Property 3 Conditional lower bound, i.e., $P_\sigma(k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j) > \min(P_\sigma(\boldsymbol{s}_i), P_\sigma(\boldsymbol{s}_j))$, $\forall k_1, k_2 \neq 0$, if only $3 \times \min(P_\sigma(\boldsymbol{s}_i), P_\sigma(\boldsymbol{s}_j)) < \max(P_\sigma(\boldsymbol{s}_i), P_\sigma(\boldsymbol{s}_j))$, where $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ denote irrelevant variables.

Proof Similar to the train of thought in upper bound, we need to prove

$$\frac{\mathbb{D}(V_{k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j}(t))}{\mathbb{E}^2(V_{k_1\boldsymbol{s}_i + k_2\boldsymbol{s}_j}(t))} > \frac{\mathbb{D}(V_{k_1\boldsymbol{s}_i}(t))}{\mathbb{E}^2(V_{k_1\boldsymbol{s}_i}(t))}. \tag{8}$$

Unfolding Eq. (8) yields

$$\frac{\mathbb{D}(Y)}{\mathbb{E}^2(Y)} > \frac{\mathbb{D}(X)}{\mathbb{E}^2(X)} + \frac{2\mathbb{D}(X)}{\mathbb{E}(X)\mathbb{E}(Y)}. \tag{9}$$

Thus, the constraint of inequality (9) would enable Eq. (8) to hold. To make Eq. (9) more intuitive, we adopt the approximation trick

$$\frac{\mathbb{D}(X)}{\mathbb{E}^2(X)} + \frac{2\mathbb{D}(X)}{\mathbb{E}(X)\mathbb{E}(Y)} \approx \frac{3 \times \mathbb{D}(X)}{\mathbb{E}^2(X)}.$$

Therefore, $3 \times \min(P_\sigma(\boldsymbol{s}_i), P_\sigma(\boldsymbol{s}_j)) < \max(P_\sigma(\boldsymbol{s}_i), P_\sigma(\boldsymbol{s}_j))$ can be a subsidiary judgement on whether the lower bound exists.

Compared with kurtosis in the traditional BSS algorithms, what notable in PE is that, in some scenarios, its extremums locate both in the upper and lower bound of the

optimizing domain, while those of kurtosis only correspond to a minimum. To conceptualize this, we adopt two Logistic signals s_1 and s_2 to compare the objective surfaces of kurt$(w_1 s_1 + w_2 s_2)$ and $P_5(w_1 s_1 + w_2 s_2)$, where $P_5(\cdot)$ means the control factor of PE is $\sigma = 5$. In this instance, the Logistic map is $s(t+1) = us(t)(1-s(t))$, where $u = 3.9$ and $u = 3.99$ are used for s_1 and s_2, respectively. As a consequence, figure 1 shows the stagnation points of PE are all valuable. On the contrary, interfering maximums exist in kurtosis. It should be emphasized here that the establishment of Property 2 is rigourous and sufficient enough for designing its BSS algorithm, and the hypothetical point of lower bound would degenerate to a local maximum if $3 \times \min(P_\sigma(s_i), P_\sigma(s_j)) \geq \max(P_\sigma(s_i), P_\sigma(s_j))$.

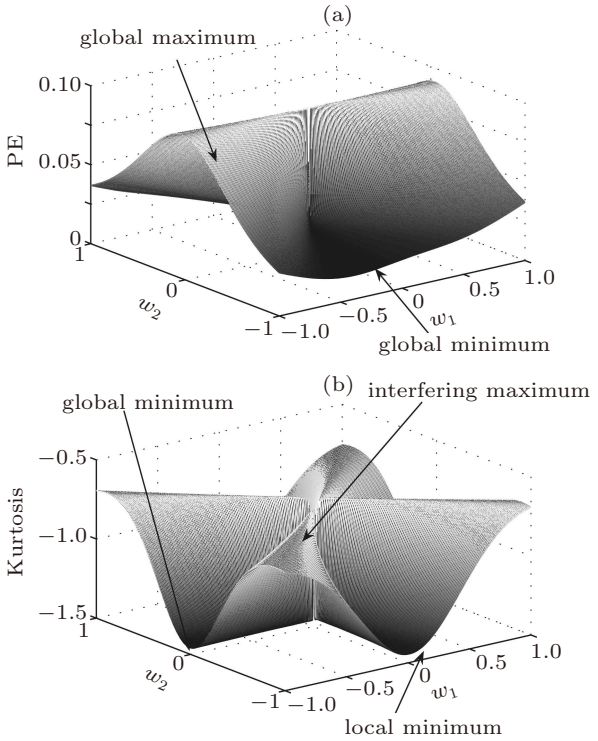


Fig. 1. Objective surfaces of (a) PE and (b) kurtosis.

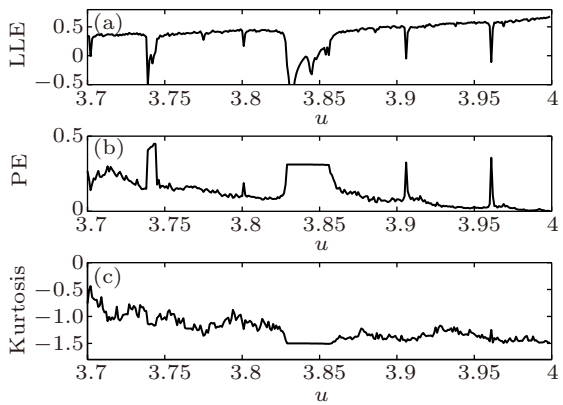


Fig. 2. (a) Fitting LLE of logistic map by (b) PE and (c) kurtosis.

PE also fits the largest Lyaponov exponent (LLE) quite consistently. For the purpose of showing a large range of LLE spectrum, a Logistic map is stilled employed in this example, while its combined rate u varies from 3.7 to 4. Figure (2) reveals that PE features almost a reverse trend of LLE except the negative LLE part. By contrast, kurtosis fails to fit LLE.

3. Blind separation algorithm

In the linear mixed model of BSS, we assume M signal arrays $\boldsymbol{S} = [s_1, s_2, \ldots, s_M]^{\mathrm{T}}$ are observed through M sensors, where the observed signals are denoted as $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M]^{\mathrm{T}}$. Thus, the BSS model can be described by the following equation:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}, \tag{10}$$

where $\boldsymbol{A} \in \mathbb{R}^{M\times M}$ is a mixing matrix, $\boldsymbol{X} \in \mathbb{R}^{M\times N}$, and $\boldsymbol{S} \in \mathbb{R}^{M\times N}$.

The mixed signals can be separated by using specific separating algorithms if the mixing matrix $\boldsymbol{A}$ has full-column rank. Suppose $\boldsymbol{w}_i$ is a separating vector where $\boldsymbol{A}^{-1}(:,i)$ is its ideal value, which suggests that a separated signal $\hat{s}_i$ can be acquired by

$$\hat{s}_i = \boldsymbol{w}_i{}^{\mathrm{T}}\boldsymbol{X}, \quad i = 1,2,\ldots,M. \tag{11}$$

Accordingly, equation (11) is transformed to a multivariate optimization problem. For the purpose of maximizing or minimizing the objective function value of $P_\sigma(\boldsymbol{w}_i{}^{\mathrm{T}}\boldsymbol{X})$, the most classic approach is to update the separating vector in a gradient direction.

Let $\boldsymbol{C}_t \triangleq (\boldsymbol{X}(:,1:d) - \boldsymbol{X}(:,1+t:d+t))(\boldsymbol{X}(:,1:d) - \boldsymbol{X}(:,1+t:d+t))^{\mathrm{T}}$, $\bar{\boldsymbol{C}} \triangleq \frac{1}{\sigma}\sum_{t=1}^{\sigma}\boldsymbol{C}_t$ and $\alpha \triangleq V_{\boldsymbol{w}_i{}^{\mathrm{T}}\boldsymbol{X}}(t), = \boldsymbol{w}_i{}^{\mathrm{T}}\boldsymbol{C}_t\boldsymbol{w}_i$ we have

$$\mathbb{E}^2(\alpha) = (\boldsymbol{w}_i{}^{\mathrm{T}}\bar{\boldsymbol{C}}\boldsymbol{w}_i)^2. \tag{12}$$

Since $\bar{\boldsymbol{C}}$ is a symmetric matrix, the derivative of $\mathbb{E}^2(\alpha)$ with respect to $\boldsymbol{w}_i$ is

$$\nabla_{\boldsymbol{w}_i}\mathbb{E}^2(\alpha) = 4\boldsymbol{w}_i{}^{\mathrm{T}}\bar{\boldsymbol{C}}\boldsymbol{w}_i\bar{\boldsymbol{C}}\boldsymbol{w}_i. \tag{13}$$

Similarly,

$$\mathbb{D}(\alpha) = \frac{1}{\sigma}\sum_{t=1}^{\sigma}(\boldsymbol{w}_i{}^{\mathrm{T}}(\boldsymbol{C}_t - \bar{\boldsymbol{C}})\boldsymbol{w}_i)^2, \tag{14}$$

and the derivative of $\mathbb{D}(\alpha)$ with respect to $\boldsymbol{w}_i$ is

$$\nabla_{\boldsymbol{w}_i}\mathbb{D}(\alpha) = \frac{4}{\sigma}\sum_{t=1}^{\sigma}\boldsymbol{w}_i{}^{\mathrm{T}}(\boldsymbol{C}_t - \bar{\boldsymbol{C}})\boldsymbol{w}_i(\boldsymbol{C}_t - \bar{\boldsymbol{C}})\boldsymbol{w}_i. \tag{15}$$

By associating Eqs. (12)–(15), we can obtain the gradient of the objective function, i.e.,

$$\nabla_{\boldsymbol{w}_i}P_\sigma(\boldsymbol{w}_i{}^{\mathrm{T}}X) = \frac{\mathbb{E}^2(\alpha)\nabla_{\boldsymbol{w}_i}\mathbb{D}(\alpha) - \mathbb{D}(\alpha)\nabla_{\boldsymbol{w}_i}\mathbb{E}^2(\alpha)}{\mathbb{E}^4(\alpha)}. \tag{16}$$

This immediately gives the following algorithm for maximal PE estimation:

$$\boldsymbol{w}_i(k) = \boldsymbol{w}_i(k-1) + \mu(k-1)\nabla_{\boldsymbol{w}_i} P_\sigma(\boldsymbol{w}_i{}^{\mathrm{T}} X), \tag{17}$$

where $\mu(k-1)$ is called learning rate.

Supposing $P_\sigma(\boldsymbol{s}_1) > \cdots > P_\sigma(\boldsymbol{s}_{M-1}) > P_\sigma(\boldsymbol{s}_M)$, the gradient ascent method of Eq. (17) would always make sense geometrically since Property 2 guarantees that $P_\sigma(\boldsymbol{s}_1), \ldots, P_\sigma(\boldsymbol{s}_{M-1})$ all correspond to local maximums when applying the orthogonalization constraint. As for $\boldsymbol{s}_M$, moving in the direction of the negative or positive gradient which is determined by its initial position would find its extremum $P_\sigma(\boldsymbol{s}_M)$. Yet in actual applications, the separating vectors of the $M-1$ maximums would overwhelmingly help us extract the correct separating vector of the Mth extremum when $M \geqslant 3$. In other words, the prejudgement of whether $P_\sigma(\boldsymbol{s}_{M-1}) > 3P_\sigma(\boldsymbol{s}_M)$ can be omitted when devising $\mu(k-1)$. Based on the above analysis, a decreasing learning rate for $\boldsymbol{w}_i(k)$ is given by

$$\mu(k-1) = \frac{2\beta}{\beta + k - 1}, \tag{18}$$

with being β an appropriate positive constant (e.g., $\beta = 100$).

Before deploying a gradient search, we first assume that both the mixture variables and the source components have zero mean without loss of generality, otherwise a centering operation should be used for pre-processing as the upper bound and conditional lower bound are built on zero-mean variables which means $\mathbb{D}(\boldsymbol{s}(t+1) - \boldsymbol{s}(1)) = \sigma||\boldsymbol{s}(t+1) - \boldsymbol{s}(1)||^2$.

Also, pre-whitening operation is encouraged because it eliminates the second order correlation between the variables and minus the convergence rounds required. Specifically, whitening means that we linearly transform the observed data vector $\boldsymbol{X}$ by linearly multiplying it with a matrix $\boldsymbol{V}$, namely,

$$\boldsymbol{X}' = \boldsymbol{V}\boldsymbol{X}, \tag{19}$$

where $\boldsymbol{V} = (\frac{1}{N}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1/2}$. Suppose the eigenvalue decomposition (EVD) of the covariance matrix is $\frac{1}{N}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = \boldsymbol{D}\boldsymbol{E}\boldsymbol{D}^{\mathrm{T}}$, thus $(\frac{1}{N}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1/2} = \boldsymbol{D}\boldsymbol{E}^{-1/2}\boldsymbol{D}^{\mathrm{T}}$.

In addition, the separating matrix $\boldsymbol{W}(k) = [\boldsymbol{w}_1(k), \ldots, \boldsymbol{w}_M(k)]$ needs to be orthogonalized after each iteration, especially in the case of separating several chaotic components, by using

$$\boldsymbol{W}(k) \leftarrow \boldsymbol{W}(k)(\boldsymbol{W}(k)^{\mathrm{T}}\boldsymbol{W}(k))^{-1/2}, \tag{20}$$

where $(\boldsymbol{W}(k)^{\mathrm{T}}\boldsymbol{W}(k))^{-1/2}$ is also obtained from an eigenvalue decomposition.

The final form of our gradient search algorithm, therefore, can be outlined as follows.

Initialization Choose the initial value of control factor σ, specify the learning rate $\mu(k-1)$ with β, and select a dimension d of the reconstructed phase space.

Preprocessing Center the input data $\boldsymbol{X}$, and write them by $\boldsymbol{X}' = \boldsymbol{V}\boldsymbol{X}$.

Updating Update all the separating vectors by

$$\boldsymbol{w}_i(k) = \boldsymbol{w}_i(k-1) + \mu(k-1)\nabla_{\boldsymbol{w}_i} P_\sigma(\boldsymbol{w}_i{}^{\mathrm{T}} X'), \quad i = 1, 2, \ldots, M. \tag{21}$$

After that, orthogonalize the matrix $\boldsymbol{W}(k) = [\boldsymbol{w}_1(k), \ldots, \boldsymbol{w}_M(k)]$ by using Eq. (20) and retrieve the renewed $\boldsymbol{W}(k)$ to $\boldsymbol{w}_1(k), \ldots, \boldsymbol{w}_M(k)$.

If not converged, continue updating. Otherwise, export the separating matrix $\boldsymbol{W}$ and separated signals $\hat{\boldsymbol{S}}$ by $\hat{\boldsymbol{S}} = \boldsymbol{W}\boldsymbol{X}'$.

4. Simulation

In this section, we evaluate the performance of our PE-based blind separation algorithm by computer simulations. Three kinds of chaotic signals are selected as the source signals. The first one is generated by Rossler system, namely,

$$\begin{aligned} \dot{\phi}_1 &= -(\phi_2 + \phi_3), \\ \dot{\phi}_2 &= \phi_1 + a\phi_2, \\ \dot{\phi}_3 &= b + \phi_3(\phi_1 - c), \end{aligned} \tag{22}$$

where $a = 0.2$, $b = 0.2$, and $c = 5$. The second one is generated by Duffing system, i.e.,

$$\begin{aligned} \dot{\varphi}_1 &= \varphi_2, \\ \dot{\varphi}_2 &= -\delta\varphi_2 - a\varphi_1(1 + {\varphi_1}^2) + b \times \cos(\varphi_3), \\ \dot{\varphi}_3 &= c, \end{aligned} \tag{23}$$

where $\delta = 0.05$, $a = 0.5$, $b = 7.5$, and $c = 1$. The last one is generated by Lorenz system, i.e.,

$$\begin{aligned} \dot{\psi}_1 &= -a(\psi_1 - \psi_2), \\ \dot{\psi}_2 &= -\psi_1\psi_3 + c\psi_1 - \psi_2, \\ \dot{\psi}_3 &= \psi_1\psi_2 - b\psi_3, \end{aligned} \tag{24}$$

where $a = 28$, $b = 8/3$, and $c = 10$.

Equations (22)–(24) are numerically integrated by the fourth-order Runge-Kutta algorithm herein, while the computing step length for Rossler, Duffing and Lorez systems are 0.05, 0.05, and 0.01, respectively. The ϕ_2, φ_2, and ψ_2 components of these systems are adopted to be the signal sources (i.e., $\boldsymbol{s}_1$, $\boldsymbol{s}_2$, and $\boldsymbol{s}_3$), and the length of each time series is $N = 4000$. The 3×3 mixing matrix $\boldsymbol{A}$ is randomly generated from the standard normal distribution in each trial. The

statistical performance or accuracy is measured using a performance index (PI)[13] defined as

$$\mathrm{PI}=\sum_{i=1}^{M}\left(\sum_{j=1}^{M}\frac{|z_{ij}|^2}{\max_k|z_{ik}|^2}-1\right)+\sum_{j=1}^{M}\left(\sum_{i=1}^{M}\frac{|z_{ij}|^2}{\max_k|z_{kj}|^2}-1\right), \quad (25)$$

where z_{ij} is the ij-th element of the matrix $\boldsymbol{Z}=\boldsymbol{WVA}$. The larger the value PI is, the poorer the statistical performance of a separation algorithm.

4.1. Blind separation in noiseless case

Since the observables are clean, the main parameters in our algorithm are set as: $\sigma=5$, $d=2000$, and $\beta=100$. For the signal model $\boldsymbol{X}=\boldsymbol{AS}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3]^{\mathrm{T}}$, the mixed signals are shown in Fig. 3. After specific iterations using gradient search, the separated signals are shown in Fig. 4, which reveals that the separated signals are quite consistent with the source signals, despite they are recovered with different orders. In addition, the PI versus iteration relationship is also presented in Fig. 5. The performance index plummets to 0.02 only after 20 iterations and levels off in the later phases. The results indicate that our algorithm can separate the mixed signals accurately at a satisfactory pace.

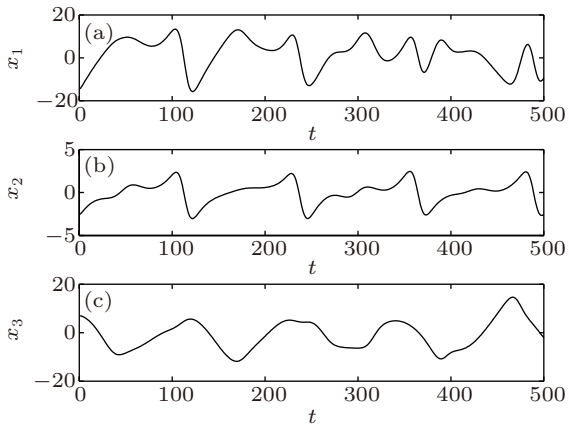


Fig. 3. Randomly mixed signals for (a) x_1, (b) x_2, and (c) x_3.

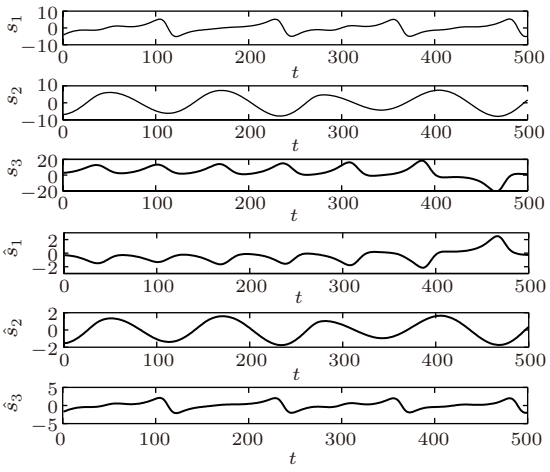


Fig. 4. Comparison between source signals and separated signals.

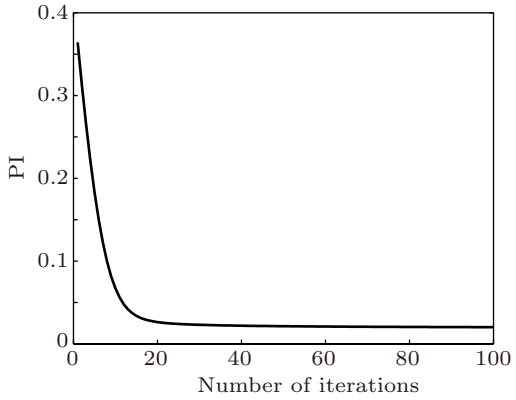


Fig. 5. The convergence attribute of PI.

4.2. Blind separation under noise contamination

When the observed signals are contaminated by noise, the robustness of our algorithm is investigated in this part. Consider the signal model $\boldsymbol{X}=\boldsymbol{AS}+\boldsymbol{N}$, where the entries of $\boldsymbol{N}$ denotes independent and identically distributed zero-mean white Gaussian noises. In the previous literatures, the covariance matrix of the noise $\boldsymbol{N}$ is often assumed to be in the form $\delta^2\boldsymbol{I}$ with known δ^2 and little work on estimation of an unknown noise covariance has been conducted.[16,17] Yet we have no transcendental knowledge of actual physical noise presents in the measuring devices especially in the context of blind signal separation. Presumably, the noise term would blur the pre-whiten matrix $\boldsymbol{V}$ and thus the separating vectors would be nonorthogonal. Yet this adverse impact on our algorithm is less severe than the traditional methods such as FastICA.

Let $\tilde{\boldsymbol{x}}_i$ denote one blurred signal, its synthetic signal to noise ratio (SNR) is calculated by

$$\mathrm{SNR}=10\lg\frac{||\boldsymbol{x}_i||^2}{||\tilde{\boldsymbol{x}}_i-\boldsymbol{x}_i||^2}, \quad i=1,2,3. \quad (26)$$

The $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ are manipulated with concurrent SNRs and adopted as the polluted mixtures. We take $\sigma=200$, $d=2000$, and $\beta=100$ in this section, along with 200 Monto Carlo trials utilized to obtain a averaged relevance of SNR versus PI. Among them, σ is set much larger than the value of the noise-free case in order to attenuate the noise effectively. The performance comparison between our method and FastICA is shown in Fig. 6. Due to the fact that FastICA and the proposed method both lose efficacy and any post-processing technique for denoising is meaningless when SNR ≤ 0 dB, we constrain the input SNR region to be 1–30 dB herein.

According to Fig. 6, firstly, there is a negative correlation between SNR and PI in both FastICA and the proposed method. The cause is that $\boldsymbol{Z}=\boldsymbol{WVA}=\boldsymbol{W}[\boldsymbol{AS}(\boldsymbol{AS})^{\mathrm{T}}+\delta^2\boldsymbol{I}]^{-1/2}\boldsymbol{A}$, so $\boldsymbol{W}[\boldsymbol{AS}(\boldsymbol{AS})^{\mathrm{T}}+\delta^2\boldsymbol{I}]^{-1/2}\boldsymbol{A}$ tends to be an orthogonal matrix when δ^2 drops, and PI would be smaller. Secondly, figure 6 shows that our method outperforms FastICA

in the entire input SNR region of 1–30 dB. It is understandable because the objective surface of FastICA which based on kurtosis has more stagnation points. The solutions of FastICA which searches $\partial \text{kurt}(\boldsymbol{w}_i^{\mathrm{T}} X)/(\partial w_i) = 0$ would lose effectiveness when the noise-blurred pre-whitening matrix renders their orthogonal constraints invalid. On the contrary, the objective surface defined by proliferation exponent has less blurring extremums, so the recovered PI value by our method is still nice under noise contamination. When the above differences reflect on programmes, one fact to note is that we can employ $||\boldsymbol{w}_i(k) - \boldsymbol{w}_i(k-1)|| < 10^{-3}$ as the termination principle of our algorithm while noisy FastICA needs to work under pre-determined number of iterations.

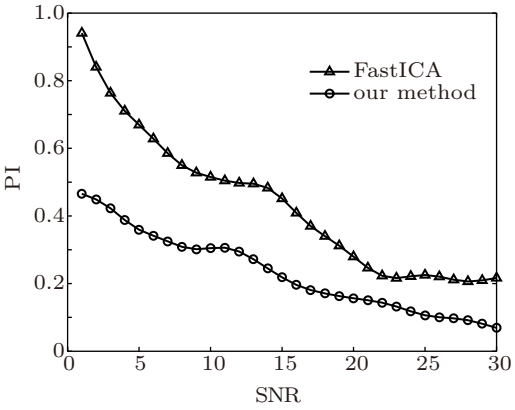


Fig. 6. Performance comparison between our method and FastICA.

4.3. Extracting one component under noise contamination

If the pre-whitening operation was omitted, which would lead to faint separating matrix, we further investigate the performance of our algorithm in the regard of extracting one component in a noisy environment by setting the parameters as $\sigma = 200$, $d = 2000$, $\beta = 100$, and 200 Monto Carlo trials are utilized to attain a averaged result.

Only the case of SNR = 10 dB is considered herein because treating other SNRs is analogous. Specifically, the absolute value of correlation coefficient $|\xi|$[10] is adopted to quantitatively evaluate the accuracy of the extracted result. In addition, $|\xi|_{\hat{s}_1,s_1}$, $|\xi|_{\hat{s}_1,s_2}$ and $|\xi|_{\hat{s}_1,s_3}$ are sorted in descend order in each trial. Theoretically, if $\hat{s}_1$ is the result of s_1 plus 10 dB noise without any aliasing with s_2 and s_3, its correlation coefficient should be $|\xi| = 0.9535$. Table 1 shows the averaged $|\xi|$ of extracted $\hat{s}_1$ with regard to distinct source signals.

Comparing the results of our method and FastICA, a striking difference is that $|\xi|_{\hat{s}_1,s_1}$ in the first row is larger than that of the second row, while $|\xi|_{\hat{s}_1,s_2}$ and $|\xi|_{\hat{s}_1,s_3}$ in the first row are smaller than those of the second row. Its explanation is, as we have mentioned previously, less interfering extremums exist in the newly defined PE. Results in Table 1 confirm the proposed method would have higher probability of locating a true extremum.

Table 1. Compared results of $|\xi|$.

	$\lvert\xi\rvert_{\hat{s}_1,s_1}$	$\lvert\xi\rvert_{\hat{s}_1,s_2}$	$\lvert\xi\rvert_{\hat{s}_1,s_3}$
Our method	0.8990	0.2522	0.1111
FastICA	0.8644	0.3618	0.1328

5. Conclusion

In this paper, we have presented a method to perform BSS for chaotic signals which started from the phase space reconstruction. The simulation results show that, our method is capable of producing successful solutions with just a few iterations in the noiseless case, while the robustness of our scheme surpasses that of FastICA in terms of anti-noise effect. Additionally, it should be stressed that the validity of PE-based method holds for chaotic maps as well, albeit our implementations are mainly carried out by chaotic flows.

References

[1] Feng J C and Tse C K 2008 *Reconstruction of Chaotic Signals with Applications to Chaos-based Communications* (Beijing: Tsinghua University Press) pp. 3–20
[2] Li W L, Li S F and Li G 2012 *Chin. Phys. B* **21** 064217
[3] Luo Y L and Du M H 2013 *Chin. Phys. B* **22** 080503
[4] Arena P, Buscarino A, Fortuna L and Frasca M 2006 *Phys. Rev. E* **74** 026212
[5] Xie Z B and Feng J C 2010 *IEEE Trans. Circuits Syst. II Exp. Briefs* **57** 461
[6] Liu K, Li H, Dai X and Xu P 2005 *J. Inf. Comput. Sci.* **2** 283
[7] Hu Z H and Feng J C 2011 *Acta Phys. Sin.* **60** 070505 (in Chinese)
[8] Wang S Y and Feng J C 2012 *Acta Phys. Sin.* **61** 170508 (in Chinese)
[9] Popivanov D, Jivkova S, Stomonyakov V and Nicolova G 2005 *Signal Process.* **85** 2112
[10] Chen H B, Feng J C and Fang Y 2008 *Chin. Phys. Lett.* **25** 405
[11] Kuraya M, Uchida A, Sano S, Yoshimori S and Umeno K 2008 *Electron. Lett.* **44** 248
[12] Hyvarinen A, Karhunen J and Oja E 2001 *Independent Component Analysis* (New York: John Wiley & Sons) pp. 147–237
[13] Hyvarinen A and Oja E 1997 *Neural Comput.* **9** 1483
[14] Bell A J and Sejnowski T J 1995 *Neural Comput.* **7** 1129
[15] Yang H H and Amari S I 1997 *Neural Comput.* **9** 1457
[16] Ikeda S and Toyama K 2000 *Neural Networks* **13** 1063
[17] Ikeda S 2000 *ICA on Noisy Data: A Factor Analysis Approach* (London: Springer) pp. 201–215
[18] Takens F 1981 *Lecture Notes in Mathematics* **898** 366