
Machine Learning, 61, 105–127, 2005

DOI: 10.1007/s10994-005-1506-8

A High Order Cumulants Based Multivariate Nonlinear Blind Source Separation Method

FENG ZHANG zhangfeng@neo.tamu.edu
Fairchild Semiconductor, 82 Running Hill Road, South Portland, ME 04106, USA

Editor: Dale Schuurmans
Published online: 11 July 2005

Abstract. This article addresses the problem of identifying multiple linear and nonlinear patterns from multivariate noisy data represented by an additive model. Following the proposed nonlinear model, the blind source separation (BSS) criterion, as a function of high-order cumulants, is shown to produce a block-structured joint cumulant matrix by an orthogonal rotation. An intuitive interpretation of this criterion is to rotate the elements of whitened principal component analysis (PCA) scores such that they are as independent as possible. The resulting optimal joint cumulant matrix contains diagonal "blocks" that correspond to the linear and nonlinear patterns caused by independent sources, from which linear patterns are recognized as in linear BSS. The nonlinear patterns are identified by extracting their lower-dimensional manifolds via the principal curves method and then transforming back to the original data space. As illustrated in the experimental study, the estimated linear and nonlinear patterns will provide more accurate diagnosing of the root causes that contribute to the observed variability in multivariate manufacturing.

Keywords: blind source separation, cumulant, cumulant matrix, principal component analysis

1. Introduction

Blind source separation is a widely used method that has been intensively studied and applied over recent decades for identifying independent sources from their linear effects on the observed data (Hyvarinen, Karhunen, & Oja, 2001). In real industrial applications, however, the collected measurement data do not always fit the assumed linear BSS model, in which the data are observed as a collection of linear or nonlinear effects from multiple potential independent sources, possibly with additive noise. Therefore, as a natural extension of linear BSS, an appropriate nonlinear model and its corresponding source separation method will have broader applicability in multivariate process analysis and pattern identification. Few previous studies have been dedicated to the problem of nonlinear blind source separation and its applications in engineering fields. On the other hand, as a nonlinear method, principal curves were proposed to generalize linear PCA for nonlinear feature extraction (Hastie & Stuetzle, 1989; Chang & Ghosh, 1998). The drawback of principal curves and other related nonlinear PCA approaches is that they can only recognize the nonlinear pattern when it is caused by a single random source. Therefore, it is worth investigating a more general technique that might separate and identify multiple nonlinear patterns present in the data. We have called this the nonlinear BSS problem.

Among the contributions to nonlinear blind source separation, Burel (1992) proposed a neural network algorithm to estimate unknown parameters when the patterns are modeled as fixed nonlinear functions of some random variables. Krob and Benidir (1994) discussed the problem of polynomial mixtures and identified the nonlinear patterns using high-order moments. Self-organizing maps (SOM) were applied to quantify the nonlinear effects in observations in which the local probability density function can be factored to approximate source independence (Pajunen, Hyvarinen, & Karhunen, 1996). However, it required a huge number of neurons for good accuracy and was limited to source density functions with bounded support. To relax this limitation, Pajunen and Karhunen (1997) introduced generative topographic mapping (GTM) by imposing output distributions to match some predefined densities. In another class of nonlinear BSS problems, the volume conservation condition was assumed on restrictive nonlinear functions (Deco & Brauer, 1995). Taleb and Jutten (1999) advocated the separation of a particular nonlinear BSS problem called post-nonlinear mixtures, in which data were assumed to be a component-wise nonlinear function of a linear instantaneous mixture of sources. In the algorithm developed in Yang et al. (1998) the nonlinearities were defined to be invertible by a two-layer perceptron. Recently, Valpola et al. (2003) adopted multilayer perceptrons to parameterize the nonlinear patterns and applied Bayesian variational treatment to identify independent sources.

These analyses of the nonlinear BSS problem lack a solid theoretical background, and were solved on a model-by-model basis (Taleb, 2002). Most approaches share a common property of conditioning the nonlinear functions in the model to simplify the nonlinear effects observed in the data (Jutten & Karhunen, 2003). The intention of this article is to introduce a nonlinear BSS model and present the corresponding source separation algorithm for multiple pattern identification. We will deal with the abovementioned drawbacks in the proposed algorithm, which: (1) automatically selects the number of unknown independent sources; (2) provides theoretical analysis of the separability of multiple linear and nonlinear patterns; and (3) avoids requirements on derivatives of nonlinear functions or predefined densities of sources.

This article is organized into six sections. Following this introduction, a nonlinear BSS model is introduced to represent multiple linear and nonlinear patterns present in multivariate data, where each nonlinear pattern is assumed to lie in a lower-dimensional feature space. Section 3 first presents a brief review of linear BSS results, and proposes a high-order cumulants-based source separation criterion. The optimal joint cumulant matrix produced by the criterion has a diagonal structure and forms the basis of the nonlinear BSS algorithm. A statistical testing method is developed in Section 4 to separate diagonal blocks automatically and to identify multiple patterns from the sample estimate of the joint cumulant matrix. Applications of the proposed algorithm are illustrated in Section 5. Finally, some discussion and ideas for future work concludes the article.

2. Nonlinear BSS model

Let $\boldsymbol{x} = [x_1, x_2, \ldots, x_d]^{\mathrm{T}}$ be the observed data from sensors and $\boldsymbol{v} = [v_1, v_2, \ldots, v_p]^{\mathrm{T}}$ the p independent sources. Consider an instantaneous nonlinear BSS problem in which $\boldsymbol{x}$ is a

general function of p unknown sources, i.e.,

$$x_i = f_i(v_1, v_2, \ldots, v_p), \qquad i = 1, 2, \ldots, d, \tag{1}$$

where x_i is the ith coordinate of $\boldsymbol{x}$. Recall that for linear BSS the source separation consists of finding a separating matrix such that the outputs are independent up to a permutation. Unlike the linear case, however, the assumption of source independence alone is not sufficient to recover $\boldsymbol{v}$ and the corresponding nonlinear functions f_i in model (1) (Taleb & Jutten, 1999; Hyvarinen & Pajunen, 1999). In other words, it is not possible to restore v_i and f_i in the general model (1) without imposing more constraints.

2.1. *Post-nonlinear mixture model*

Taleb and Jutten (1999) proposed a nonlinear BSS algorithm for the post-nonlinear mixture model as below:

$$x_i = f_i\left(\boldsymbol{c}_i^{\mathrm{T}}\boldsymbol{v}\right), \qquad i = 1, 2, \ldots, d, \tag{2}$$

where the functions f_i are required to be differentiable and invertible, and matrix $\mathbf{C} = [\boldsymbol{c}_1\boldsymbol{c}_2 \ldots \boldsymbol{c}_d]$ should be regular such that there are at least two nonzero entries in each row or column. Moreover, for simplicity, the noiseless model (2) is restricted to the square case when $d = p$.

In real engineering applications, however, the number of sensors, d, may be much greater than that of the underlying sources, i.e., $d \gg p$. Another possible difficulty for the implementation of the above method arises from the differentiability requirement on f_i. As discussed later in Section 5, nonlinear patterns f_i are not always differentiable (e.g., piecewise linear curves). To relax these restrictions on the nonlinear BSS problem, a new model is proposed in this article with weaker assumptions on functions f_i and matrix $\mathbf{C}$, which also takes into consideration the effects of measurement noise.

2.2. *An additive nonlinear BSS nodel*

In the proposed nonlinear BSS model (3), v_i is the ith unit-variance random source, which has a linear effect on $\boldsymbol{x}$ characterized by a vector $\boldsymbol{c}_i$ ($i = 1, 2, \ldots, p$), and $\boldsymbol{f}_j(t_j)$ ($j = 1, 2, \ldots, q$) is a zero-mean nonlinear pattern caused by sources t_j. All the sources v_i and t_j are assumed to be statistically independent, and independent of noise vector $\boldsymbol{\varepsilon}$,

$$\begin{aligned}\boldsymbol{x} &= v_1\boldsymbol{c}_1 + v_2\boldsymbol{c}_2 + \cdots + v_p\boldsymbol{c}_p + \boldsymbol{f}_1(t_1) + \boldsymbol{f}_2(t_2) + \cdots + \boldsymbol{f}_q(t_q) + \boldsymbol{\varepsilon} \\ &= [\boldsymbol{c}_1\boldsymbol{c}_2 \ldots \boldsymbol{c}_p][v_1\, v_2 \cdots v_p]^{\mathrm{T}} + \sum_{j=1}^{q} \boldsymbol{f}_j(t_j) + \boldsymbol{\varepsilon} \\ &\equiv \mathbf{C}\boldsymbol{v} + \sum_{j=1}^{q} \boldsymbol{f}_j(t_j) + \boldsymbol{\varepsilon}. \end{aligned} \tag{3}$$

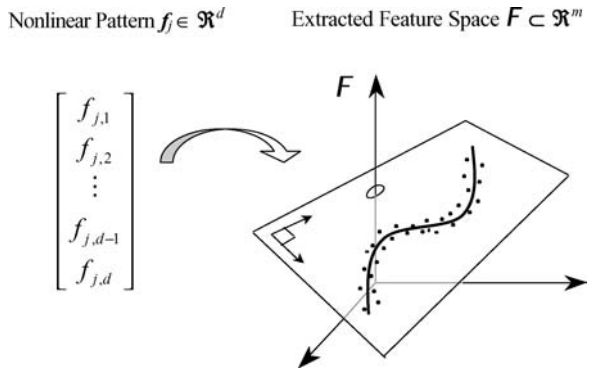

Figure 1. The nonlinear pattern $\boldsymbol{f}_j$ can be recognized by the principal curves method when it exhibits a lower-dimensional feature embedded in the d-dimensional data space.

Noise $\boldsymbol{\varepsilon}$ is a $d \times 1$ zero-mean random vector, representing the aggregated effects of measurement noise and any inherent unmodeled variation. Throughout this article, it is assumed that the covariance matrix of $\boldsymbol{\varepsilon}$ is $\boldsymbol{\Sigma}_\varepsilon = \sigma^2 \mathbf{I}$. The goal of the nonlinear BSS algorithm proposed subsequently is to separate the linear and nonlinear effects in $\boldsymbol{x}$ by identifying every linear pattern vector $\boldsymbol{c}_i$ and nonlinear pattern $\boldsymbol{f}_j$.

In advanced modern manufacturing, for example, it is not uncommon to have multiple independent root causes or sources that contribute to the observed variability in measurement data. Each source will result in a distinct spatial pattern across some of the measured variables in $\boldsymbol{x}$, indicating how the source causes them to interact through a linear or nonlinear relationship. Often such effects from unknown sources are applied to data $\boldsymbol{x}$ in an additive manner. Therefore, model (3) is an appropriate way to represent the potential linear and nonlinear effects by $\boldsymbol{c}_i$ and $\boldsymbol{f}_j$ respectively.

As demonstrated in a real example in Section 5, the multivariate nonlinear pattern $\boldsymbol{f}_j$ always exhibits a lower-dimensional feature embedded in the d-dimensional data space. The implication that the signal components of $\boldsymbol{f}_i$ lie in a lower-dimensional feature space allows us to extract the nonlinear features of $\boldsymbol{f}_i$ by the principal curves method, as shown in Figure 1.

By definition, a principal curve $\boldsymbol{g} \in \Re^m$ is a one-dimensional smooth curve that passes through the middle of the m-dimensional data $\boldsymbol{y}$, i.e.,

$$\boldsymbol{g}(t) = [g_1(t)\ g_2(t) \ldots g_m(t)]^{\mathrm{T}} = \mathbf{E}[\boldsymbol{y} \mid t_g(\boldsymbol{y}) = t],$$

where the projection index $t_g(\boldsymbol{y}) = \sup_u\{t : \|\boldsymbol{y} - \boldsymbol{g}(u)\| = \inf_v \|\boldsymbol{y} - \boldsymbol{g}(v)\|\}$. That is, each point on the principal curve is the average of all points projecting onto it. As an extension of PCA, the principal curve $\boldsymbol{g}$ characterizes the nonlinearities in data $\boldsymbol{y}$.

Note that principal curves are free of parametric forms and are defined by an ordered (via t) list of points in $\Re^m$. Thus, the nonparametric form of $\boldsymbol{f}_j$ in model (3) is flexible enough to model a variety of nonlinear patterns that may not be differentiable.

Given model (3), the proposed nonlinear BSS algorithm starts with a whiten step that transforms $\boldsymbol{x} \in \Re^d$ to uncorrelated PCA scores. The PCA scores are then rotated by an orthogonal matrix such that they are partitioned into independent groups. The final step of the algorithm is to separate the groups and identify their corresponding linear or nonlinear patterns individually.

2.3. Principal curves classification model

The form of model (3) is similar to what is assumed in the principal curves classification method (Chang & Ghosh, 1998). Given $\boldsymbol{x}$ and its covariance matrix $\boldsymbol{\Sigma}_x$, PCA decomposes $\boldsymbol{x}$ into a linear combination of eigenvectors $\boldsymbol{e}_i$ in descending order of PCA scores s_i $\{i = 1, 2, \ldots, d\}$, i.e.,

$$\begin{aligned}\boldsymbol{x} &= [s_1\boldsymbol{e}_1 + s_2\boldsymbol{e}_2 + \cdots s_k\boldsymbol{e}_k] + \cdots + [s_{d-k+1}\boldsymbol{e}_{d-k+1} + \cdots + s_d\boldsymbol{e}_d] \\ &\approx [f_{1,1}(t_1)\boldsymbol{e}_1 + f_{1,2}(t_1)\boldsymbol{e}_2 + \cdots + f_{1,k}(t_1)\boldsymbol{e}_k] + \cdots + [f_{l,1}(t_l)\boldsymbol{e}_{d-k+1} \\ &\quad + \cdots + f_{l,k}(t_l)\boldsymbol{e}_d] \\ &= \boldsymbol{f}_1(t_1) + \cdots + \boldsymbol{f}_l(t_l), \end{aligned} \tag{4}$$

where t_j is the projection index for the jth principal curve $\boldsymbol{f}_j$ and $f_{j,m}$ is the mth coordinate element ($1 \le j \le \cdots \le l = d/k$). If d is not a multiple of k, the dimension of the last principal curve is set to be the remainder of d/k.

One obvious disadvantage of this method for nonlinear pattern recognition is that it does not account for the independence between different t_i's. Although model (4) seeks to discover the nonlinearities in $\boldsymbol{x}$ by extracting multiple principal curves, it yields a somewhat artificial interpretation by predefining the dimension of every principal curve such that the dimensions of all principal curves are the same.

To accommodate these drawbacks, we introduce the nonlinear model (3) and provide a generic and black-box (i.e., requiring less prior knowledge and restrictions) means of uniquely identifying the nonlinear features present in the measurement data.

3. Nonlinear blind source separation algorithm

3.1. Review on linear BSS

To derive our nonlinear BSS algorithm, we first present a brief review of higher-order cumulants-based linear BSS methods. For the linear model $\boldsymbol{x}_{d\times 1} = \mathbf{C}_{d\times p}\boldsymbol{v}_{p\times 1} + \varepsilon$ with the same assumptions as in model (3), PCA is applied as a whitening step to $\boldsymbol{x}$ for dimension reduction. Let $\{z_k, \lambda_{x,k}\}_{k=1}^d$ be the eigenvectors and eigenvalues of $\boldsymbol{\Sigma}_x$, arranged in the order

$$\lambda_{x,1} \ge \lambda_{x,2} \ge \cdots \ge \lambda_{x,p} > \sigma^2 = \lambda_{x,p+1} = \cdots = \lambda_{x,d}.$$

Denote $\mathbf{Z}_p = [z_1, z_2, \ldots, z_p]$ and $\boldsymbol{\Lambda}_p = \mathrm{Diag}\{\lambda_{x,1}, \lambda_{x,2}, \ldots, \lambda_{x,p}\}$. Then the reduced p-dimensional PCA score vector $\boldsymbol{y} = \mathbf{W}^{-1}\boldsymbol{x}$ is transformed by the whitening matrix $\mathbf{W}^{-1}$

$(\mathbf{\Lambda}_p - \sigma^2\mathbf{I})^{-1/2}\,\mathbf{Z}_p^{\mathrm{T}}$. Working on the reduced $\boldsymbol{y}$ rather than on $\boldsymbol{x}$, linear BSS methods utilize the properties of high-order cumulants to estimate the matrix $\mathbf{C}$.

For an arbitrary zero-mean random vector $\boldsymbol{s} = [s_1, \ldots, s_m]^{\mathrm{T}}$ with finite fourth-order moments, its fourth-order cumulant is

$$\begin{aligned}\mathrm{Cum}_{i,j,k,l}(\boldsymbol{s}) = {} & E[s_i s_j s_k s_l] - E[s_i s_j]E[s_k s_l] - E[s_i s_k]E[s_j s_l] \\ & - E[s_i s_l]E[s_j s_k] \quad (1 \le i, j, k, l \le m).\end{aligned}$$

In particular, the kurtosis of s_i is the fourth-order autocumulant, i.e., $k(s_i) \equiv \mathrm{Cum}_{i,i,i,i}(\boldsymbol{s}) = E[s_i^4] - 3E^2[s_i^2]$. A cumulant involving different variables is called a cross-cumulant. Note that cross-cumulants $\mathrm{Cum}_{i,j,k,l}(\boldsymbol{s})$ $(i, j, k, l \neq i, i, i, i)$ are zero if the elements of $\boldsymbol{s}$ are mutually independent. The $m \times m$ cumulant matrix $\mathbf{Q}_s(\mathbf{M})$ of $\boldsymbol{s}$ is defined component-wise (Cardoso, 1998)

$$[\mathbf{Q}_s(\mathbf{M})]_{ij} = \sum_{k,l=1}^{m} \mathrm{Cum}_{i,j,k,l}(\boldsymbol{s})\mathrm{M}_{kl},$$

where $\mathbf{M}$ is an arbitrary $m \times m$ matrix.

For the linear BSS model, it is straightforward to establish the following structure of $\mathbf{Q}_{\mathrm{X}}(\mathbf{M})$ by cumulant properties (Cardoso, 1998):

$$\mathbf{Q}_{\mathrm{X}}(\mathbf{M}) = \mathbf{C}\mathbf{\Delta}(\mathbf{M})\mathbf{C}^{\mathrm{T}}, \tag{5}$$

where $\mathbf{\Delta}(\mathbf{M}) = \mathrm{Diag}(k(v_1)\boldsymbol{c}_1^{\mathrm{T}}\mathbf{M}\boldsymbol{c}_1, \ldots k(v_p)\boldsymbol{c}_p^{\mathrm{T}}\mathbf{M}\boldsymbol{c}_p)$. In the factorization of $\mathbf{Q}_{\mathrm{X}}(\mathbf{M})$ in Eq. (5), the (generally unknown) kurtosis $k(v_i)$ enters only in the diagonal matrix $\mathbf{\Delta}(\mathbf{M})$. Recall that PCA decomposition of $\mathbf{\Sigma}_x$ leads to $\mathbf{C} = \mathbf{Z}_p(\mathbf{\Lambda}_p - \sigma^2\mathbf{I})^{1/2}\mathbf{Q}$ where $\mathbf{Q} = [\boldsymbol{q}_1\boldsymbol{q}_2 \ldots \boldsymbol{q}_p]$ is a $p \times p$ orthogonal matrix (Apley & Lee, 2002). The cumulant matrix for the reduced vector $\boldsymbol{y} = \mathbf{W}^{-1}\boldsymbol{x} = \mathbf{W}^{-1}[\mathbf{C}\boldsymbol{v} + \varepsilon] = \mathbf{Q}\boldsymbol{v} + \mathbf{W}^{-1}\varepsilon$, can be written as

$$\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}) = \mathbf{Q}\tilde{\Delta}(\mathbf{M})\mathbf{Q}^{\mathrm{T}}, \quad \text{where } \tilde{\Delta}(\mathbf{M}) = \mathrm{Diag}\big(k(v_1)\boldsymbol{q}_1^{\mathrm{T}}\mathbf{M}\boldsymbol{q}_1, \ldots k(v_p)\boldsymbol{q}_p^{\mathrm{T}}\mathbf{M}\boldsymbol{q}_p\big). \tag{6}$$

As an eigen-decomposition, Eq. (6) transforms the problem of estimating $\mathbf{C}$ into the problem of finding the diagonalizer $\mathbf{Q}$ of $\mathbf{Q}_{\mathrm{Y}}(\mathbf{M})$. However, the eigenvectors $\boldsymbol{q}_i$ of cumulant matrix $\mathbf{Q}_{\mathrm{Y}}(\mathbf{M})$ are uniquely determined if and only if the eigenvalues are all distinct. If $\mathbf{M}$ is randomly chosen, then the eigenvalues are distinct with probability 1 (Cardoso, 1998). An appropriate selection of $\mathbf{M}$ requires prior knowledge about the unknown linear mixture, which is impossible in practice. A reasonable way to alleviate this problem is to process multiple cumulant matrices jointly. Denote $\mathrm{Off}(\mathbf{F}) = \sum_{i \neq j}(f_{ij})^2$ as the sum of the squares of the off-diagonal elements of $\mathbf{F}$. Let $M = \{\mathbf{M}_1, \ldots \mathbf{M}_{p^2}\}$ be a set of p^2 $p \times p$ matrices. The optimal $\mathbf{Q}$ for the linear BSS model is determined by minimizing the nonnegative joint diagonality criterion,

$$\Phi_M(\mathbf{V}) \equiv \sum_{\mathbf{M}_i \in M} \mathrm{Off}(\mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}_i)\mathbf{V}), \tag{7}$$

which measures how close to diagonality an orthonormal matrix $\mathbf{V}$ can simultaneously bring the cumulant matrices generated by M. The optimal matrix $\mathbf{V}$ resulting from Eq. (7) is taken as a robust estimate of $\mathbf{Q}$.

3.2. *The diagonal block-structured joint cumulant matrix*

As assumed in model (3), nonlinear patterns $\boldsymbol{f}_j(t_j)$ always lie in lower-dimensional subspaces of $\Re^d$. Suppose $\boldsymbol{f}_j(t_j)$ lies in an r_j-dimensional ($r_j < d$) linear variety (a translated linear subspace) for all t, and no other linear variety in which $\boldsymbol{f}_j(t_j)$ lies has dimension smaller than r_j, then the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}_{f_j}$ for $\boldsymbol{f}_j$, will have

$$\lambda_{j,1} \geq \lambda_{j,2} \geq \cdots \geq \lambda_{j,r_j} > 0 = \lambda_{j,r_j+1} = \lambda_{j,r_j+2} \cdots = \lambda_{j,d}, \tag{8}$$

Thus, each nonlinear pattern $\boldsymbol{f}_j$ can be represented by its PCA decomposition, that is,

$$\boldsymbol{f}_j(t_j) = s_{j,1}\boldsymbol{e}_{j,1} + \cdots + s_{j,r_j}\boldsymbol{e}_{j,r_j} = \mathbf{E}_j \boldsymbol{s}_j^{\mathrm{T}}, \quad j = 1, 2, \ldots, q, \tag{9}$$

where the $d \times p$ matrix $\mathbf{E}_j = [\boldsymbol{e}_{j,1}\boldsymbol{e}_{j,2}\cdots\boldsymbol{e}_{j,r_j}]$ corresponds to nonzero eigenvalues in Eq. (8), and $\boldsymbol{s}_j = [s_{j,1}s_{j,2}\ldots s_{j,r_j}]^{\mathrm{T}}$ is given by $s_{j,k} = \boldsymbol{f}_j^{\mathrm{T}}\boldsymbol{e}_{j,k}$ ($1 \leq k \leq r_j$).

PCA decompositions in Eq. (9) produce a linear form for model (3) as below:

$$\begin{aligned}\boldsymbol{x} &= \mathbf{C}\boldsymbol{v} + s_{1,1}\boldsymbol{e}_{1,1} + \cdots + s_{1,r_1}\boldsymbol{e}_{1,r_1} + \cdots + s_{q,1}\boldsymbol{e}_{q,1} + \cdots + s_{q,r_q}\boldsymbol{e}_{q,r_q} + \boldsymbol{\varepsilon} \\ &= [\mathbf{C}\mathbf{E}_1 \cdots \mathbf{E}_q]\left[\boldsymbol{v}^{\mathrm{T}}\boldsymbol{s}_1^{\mathrm{T}} \cdots \boldsymbol{s}_1^{\mathrm{T}}\right]^{\mathrm{T}} + \boldsymbol{\varepsilon} \equiv \mathbf{A}\boldsymbol{s} + \boldsymbol{\varepsilon}.\end{aligned} \tag{10}$$

Theorem. *Suppose that $\boldsymbol{f}_j(t_j)$ lies in an r_j-dimensional linear variety and that matrix $\mathbf{C}$ is of full rank p. Suppose also that the sets of eigenvectors from Σ_{f_j} and Σ_{f_l} ($j \neq l$) are orthogonal to each other. Denote $r = p + \sum_{j=1}^{q} r_j$. Then given the joint diagonality criterion* (7) *and its optimal orthonormal matrix $\mathbf{V}$, the rotated joint cumulant matrix $\sum_{\mathbf{M}_i \in M} \mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}_i)\mathbf{V}$ for the r-length whitened vector $\mathbf{y}$ will have a diagonal block structure.*

Proof: Given the assumptions on $\mathbf{C}$ and the eigenvector sets of Σ_{f_j} ($j = 1, 2, \ldots, q$), the length of $\boldsymbol{s} = [\boldsymbol{v}^{\mathrm{T}}\boldsymbol{s}_1^{\mathrm{T}}\ldots\boldsymbol{s}_q^{\mathrm{T}}]^{\mathrm{T}}$ in Eq. (10) is equal to $r = p + \sum_{j=1}^{q} r_j$, which is also the number of dominant eigenvalues that are greater than σ^2. It follows from the source assumption that $\boldsymbol{v}$ and $\boldsymbol{s}_j$ in Eq. (10) are independent. The elements $s_{j,k}$ in $\boldsymbol{s}_j$, however, are uncorrelated PCA scores and are not necessarily independent.

The covariance matrix $\boldsymbol{\Sigma}_x$ for $\boldsymbol{x}$ in Eq. (10) is

$$\begin{aligned}\boldsymbol{\Sigma}_x &= \mathbf{C}\mathbf{C}^{\mathrm{T}} + \sum_{k=1}^{r_1} \boldsymbol{e}_{1,k}\boldsymbol{e}_{1,k}^{\mathrm{T}}\mathrm{var}(s_{1,k}) + \cdots + \sum_{k=1}^{r_q} \boldsymbol{e}_{q,k}\boldsymbol{e}_{q,k}^{\mathrm{T}}\mathrm{var}(s_{q,k}) + \sigma^2\mathbf{I} \\ &= \mathbf{C}\mathbf{C}^{\mathrm{T}} + \mathbf{E}_1(\boldsymbol{\Lambda}_{f,1} - \sigma^2\mathbf{I})\mathbf{E}_1^{\mathrm{T}} + \cdots + \mathbf{E}_q(\boldsymbol{\Lambda}_{f,q} - \sigma^2\mathbf{I})\mathbf{E}_q^{\mathrm{T}} + \sigma^2\mathbf{I},\end{aligned} \tag{11}$$

where $\mathbf{\Lambda}_{f,j} = \mathrm{Diag}\{\lambda_{j,1}, \ldots, \lambda_{j,r_j}\}$. PCA decomposition provides another form of $\mathbf{\Sigma}_x$, i.e.,

$$\mathbf{\Sigma}_x = \mathbf{Z}_p(\mathbf{\Lambda}_p - \sigma^2\mathbf{I})\mathbf{Z}_p^{\mathrm{T}} + \mathbf{Z}_{r_1}\big(\mathbf{\Lambda}_{r_1} - \sigma^2\mathbf{I}\big)\mathbf{Z}_{r_1}^{\mathrm{T}} + \cdots + \mathbf{Z}_{r_q}\big(\mathbf{\Lambda}_{r_q} - \sigma^2\mathbf{I}\big)\mathbf{Z}_{r_q}^{\mathrm{T}} + \sigma^2\mathbf{I}, \tag{12}$$

where $\mathbf{Z}_p$ and $\mathbf{Z}_{r_j}$ consist of p and r_j eigenvectors of $\mathbf{\Sigma}_x$ respectively, corresponding to the dominant eigenvalues lying in the diagonal matrices $\mathbf{\Lambda}_p$ and $\mathbf{\Lambda}_{r_j}$ $(j = 1, 2, \ldots, q)$.

To make Eqs. (11) and (12) consistent, we have

$$\mathbf{C} = \mathbf{Z}_p(\mathbf{\Lambda}_p - \sigma^2\mathbf{I})^{1/2}\mathbf{Q}, \tag{13}$$

and

$$\mathbf{E}_j = \mathbf{Z}_{r_j} \quad \text{and} \quad \mathbf{\Lambda}_{f,j} = \mathbf{\Lambda}_{r_j}. \tag{14}$$

Following Eqs. (13) and (14), the whitened vector $\boldsymbol{y}$ is

$$\begin{aligned}
\boldsymbol{y} &= \mathbf{W}^{-1}\boldsymbol{x} = (\mathbf{\Lambda}_r - \sigma^2\mathbf{I})^{-1/2}\mathbf{Z}_r^{\mathrm{T}}(\mathbf{C}\boldsymbol{v} + \mathbf{E}_1\boldsymbol{s}_1 + \cdots + \mathbf{E}_q\boldsymbol{s}_q + \boldsymbol{\varepsilon}) \\
&= \begin{bmatrix} (\mathbf{\Lambda}_p - \sigma^2\mathbf{I})^{-1/2} & & & \\ & (\mathbf{\Lambda}_{r_1} - \sigma^2\mathbf{I})^{-1/2} & & \\ & & \ddots & \\ & & & \big(\mathbf{\Lambda}_{r_q} - \sigma^2\mathbf{I}\big)^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_p^{\mathrm{T}} \\ \mathbf{Z}_{r_1}^{\mathrm{T}} \\ \vdots \\ \mathbf{Z}_{r_q}^{\mathrm{T}} \end{bmatrix} \\
&\quad \times (\mathbf{C}\boldsymbol{v} + \mathbf{E}_1\boldsymbol{s}_1 + \cdots + \mathbf{E}_q\boldsymbol{s}_q + \boldsymbol{\varepsilon}) \\
&= \begin{bmatrix} \mathbf{Q}\boldsymbol{v} \\ \big(\mathbf{\Lambda}_{r_1} - \sigma^2\mathbf{I}\big)^{-1/2}\boldsymbol{s}_1 \\ \big(\mathbf{\Lambda}_{r_2} - \sigma^2\mathbf{I}\big)^{-1/2}\boldsymbol{s}_2 \\ \vdots \\ \big(\mathbf{\Lambda}_{r_q} - \sigma^2\mathbf{I}\big)^{-1/2}\boldsymbol{s}_q \end{bmatrix} + \mathbf{W}^{-1}\boldsymbol{\varepsilon} \equiv \begin{bmatrix} \mathbf{Q}\boldsymbol{v} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{\boldsymbol{s}}_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \tilde{\boldsymbol{s}}_q \end{bmatrix} + \boldsymbol{w} \\
&\equiv \boldsymbol{y}_v + \boldsymbol{y}_{s,1} + \cdots + \boldsymbol{y}_{s,q} + \boldsymbol{w},
\end{aligned} \tag{15}$$

where $\mathbf{\Lambda}_r = \mathrm{Diag}\{\mathrm{Diag}\{\mathbf{\Lambda}_p\}, \mathrm{Diag}\{\mathbf{\Lambda}_{r_1}\}, \ldots, \mathrm{Diag}\{\mathbf{\Lambda}_{r_q}\}\}$ and $\mathbf{Z}_r = [\mathbf{Z}_p\mathbf{Z}_{r_1}\ldots\mathbf{Z}_{r_q}]$. The elements of the $r_j \times 1$ vector $\tilde{\boldsymbol{s}}_j$ in Eq. (15) are scaled random variables with unit variance.

Thus, $\mathbf{Q}_{\mathrm{Y}}(\mathbf{M})$ is a sum of $r \times r$ cumulant matrices for independent $\boldsymbol{y}_v$ and $\boldsymbol{y}_{s,m}$ ($m = 1, \ldots, q$):

$$\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}) = \mathbf{Q}_{\mathrm{Y}_v}(\mathbf{M}) + \sum_{m=1}^{q} \mathbf{Q}_{\mathrm{Y}_{s,m}}(\mathbf{M}), \tag{16}$$

For notation simplicity, define

$$\begin{aligned} \mathbf{M}_{ij}^{v} &= \mathbf{M}_{ij} && 1 \le i, j \le p \quad \text{and} \\ \mathbf{M}_{ij}^{s,m} &= \mathbf{M}_{p+\sum_{n=1}^{m-1} r_n + i,\, p+\sum_{n=1}^{m-1} r_n + j} && 1 \le i, j \le r_m, \end{aligned}$$

as a $p \times p$ and an $r_m \times r_m$ matrix, respectively. The first term in Eq. (16) has the same structure as Eq. (6) in that $\mathbf{C}$ and $\boldsymbol{v}$ obey the same assumptions as in the linear BSS model,

$$\mathbf{Q}_{\mathrm{Y}_v}(\mathbf{M}) = \mathrm{Diag}\big(\mathbf{Q}\boldsymbol{\Delta}_v(\mathbf{M}^v)\mathbf{Q}^{\mathrm{T}}, \mathbf{0}_{r_1 \times r_1}, \ldots, \mathbf{0}_{r_q \times r_q}\big), \tag{17}$$

where $\boldsymbol{\Delta}_v(\mathbf{M}^v) = \mathrm{Diag}(k(v_1)\boldsymbol{q}_1^{\mathrm{T}}\mathbf{M}^v\boldsymbol{q}_1, \ldots k(v_p)\boldsymbol{q}_p^{\mathrm{T}}\mathbf{M}^v\boldsymbol{q}_p)$.

For each term $\mathbf{Q}_{\mathrm{Y}_{s,m}}(\mathbf{M})$ in Eq. (16), the zero elements in $\boldsymbol{y}_{s,m}$ cause it to be a zero matrix except for the $r_m \times r_m$ nonzero submatrix lying in the diagonal position, i.e.,

$$\mathbf{Q}_{\mathrm{Y}_{s,m}}(\mathbf{M}) = \mathrm{Diag}\big(\mathbf{0}_{p\times p}, \mathbf{0}_{r_1\times r_1}, \ldots, \mathbf{Q}_{\tilde{s}_m}(\mathbf{M}^{s,m}), \ldots, \mathbf{0}_{r_q \times r_q}\big),$$

where $[\mathbf{Q}_{\tilde{s}_m}(\mathbf{M}^{s,m})]_{i,j} = \sum_{k,l=1}^{r_m} \mathrm{Cum}_{i,j,k,l}(\tilde{\boldsymbol{s}}_m)\mathrm{M}_{kl}^{s,m}$ is an $r_m \times r_m$ cumulant matrix of the scaled random vector $\tilde{\boldsymbol{s}}_m$. The square matrix $\mathbf{Q}_{\tilde{s}_m}(\mathbf{M}^{s,m})$ is by definition a symmetric matrix, and decomposed by its eigenvectors as

$$\mathbf{Q}_{\tilde{s}_m}(\mathbf{M}^{s,m}) = \mathbf{P}_m\boldsymbol{\Delta}_{\tilde{s}_m}\mathbf{P}_m^{\mathrm{T}}, \quad m = 1, 2, \ldots, q. \tag{18}$$

Therefore, it is straightforward to establish the diagonal block structure for $\mathbf{Q}_{\mathrm{Y}}(\mathbf{M})$ using Eqs. (16)–(18):

$$\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}) = \begin{bmatrix} \mathbf{Q}\boldsymbol{\Delta}_v(\mathbf{M}^v)\mathbf{Q}^{\mathbf{T}} & & & \\ & \mathbf{P}_1\boldsymbol{\Delta}_{\tilde{s}_1}\mathbf{P}_1^{\mathrm{T}} & & \\ & & \ddots & \\ & & & \mathbf{P}_q\boldsymbol{\Delta}_{\tilde{s}_q}\mathbf{P}_q^{\mathrm{T}} \end{bmatrix}_{r\times r}. \tag{19}$$

The joint diagonality criterion (7) that aims to diagonalize each cumulant matrix $\mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}_i)\mathbf{V}$ will produce an $r \times r$ optimal orthogonal matrix with the same structure

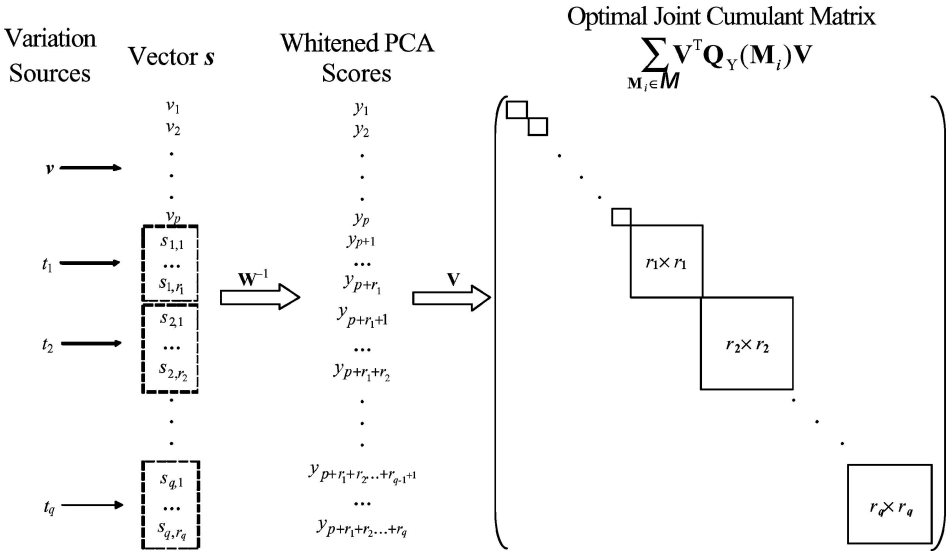

Figure 2. The diagonal block structure of joint cumulant matrix given the optimal rotation matrix $\mathbf{V}$.

as $\mathbf{Q}_{\mathrm{Y}}(\mathbf{M})$ in Eq. (19), that is,

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}^{p}_{p\times p} & & & \\ & \mathbf{V}^{s,1}_{r_1\times r_1} & & \\ & & \ddots & \\ & & & \mathbf{V}^{s,q}_{r_q\times r_q} \end{bmatrix}_{r\times r}.$$

This diagonal matrix structure minimizes the sum of squares of the off-diagonal elements of $\mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}_i)\mathbf{V}$.

Because $\mathbf{Q}$ is independent of the choice of matrix $\mathbf{M}_i$, we can always find $\mathbf{V}^{p} = \mathbf{Q}$ to achieve the diagonality of the $p \times p$ submatrix $\boldsymbol{\Delta}_v(\mathbf{M}^v)$ in Eq. (19). The eigenvector matrices $\mathbf{P}_m$, however, depend on $\mathbf{M}_i^{s,m}$ and $\mathbf{M}_i$. Thus, each $r_m \times r_m$ matrix $\mathbf{V}^{s,m}$ sought to diagonalize $\mathbf{Q}_{\tilde{s}_m}(\mathbf{M}_i^{s,m})$ in Eq. (18) is different concerning the distinct matrices $\mathbf{M}_i$ in $\boldsymbol{M}$. In other words, for the matrix set $\boldsymbol{M}$, no matrix $\mathbf{V}^{s,m}$ or $\mathbf{V}$ can achieve joint diagonality for all $\mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\tilde{s}_m}(\mathbf{M}_i)$ $\mathbf{V}$ or $\mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}_i)\mathbf{V}$.

Under criterion (7), each rotated cumulant matrix $\mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}_i)\mathbf{V}$ is not necessarily a standard diagonal matrix for the optimal matrix $\mathbf{V}$. Consequently, the sum of these rotated cumulant matrices, i.e., $\sum_{\mathbf{M}_i \in \boldsymbol{M}} \mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}_i)\mathbf{V}$, will have a block diagonal structure, as illustrated in Figure 2. □

3.3. *Multivariate pattern identification*

3.3.1. Linear pattern identification. For convenience, denote the upper $p \times p$ diagonal matrix in Figure 2 as a "linear block", and the $r_j \times r_j$ matrices along the diagonal position as "nonlinear blocks". As implied by the theoretical result in Section 3.2, these $q+1$ blocks then correspond to the linear and nonlinear patterns in model (3). For linear patterns, the joint diagonality criterion (7) is shown to produce $\mathbf{Q} = \mathbf{V}^p$. Thus, following Equation (13) we can obtain estimates of vectors $\boldsymbol{c}_i (i = 1, 2, \ldots, p)$:

$$\boldsymbol{c}_i = \mathbf{Z}_p(\boldsymbol{\Lambda}_p - \sigma^2\mathbf{I})^{1/2}\boldsymbol{q}_i \,. \tag{20}$$

In the same fashion as in linear BSS, the independent source $\boldsymbol{v}$ is recovered by

$$\boldsymbol{v} = \mathbf{V}^{p,\mathrm{T}}[y_1, \ldots, y_p]^{\mathrm{T}} \,, \tag{21}$$

where $y_1, \ldots, y_p$ are the first p elements of $\boldsymbol{y}$ corresponding to the linear block in Figure 2.

3.3.2. Nonlinear pattern identification. Nonlinear pattern identification is more complicated than the case above because we cannot directly estimate $\boldsymbol{f}_j$ using Eq. (20). Let $\boldsymbol{y}^j = [y_{p+\sum_{n=1}^{j-1} r_n+1} y_{p+\sum_{n=1}^{j-1} r_n+2} \cdots y_{p+\sum_{n=1}^{j-1} r_n+r_j}]^{\mathrm{T}}$ be an $r_j \times 1$ subvector in $\boldsymbol{y}$ that corresponds to the jth nonlinear block.

Corollary. *Consider the jth nonlinear block and its associated $r_j \times r_j$ diagonal eigenvalue matrix $\boldsymbol{\Lambda}_{r_j}$ and $d \times r_j$ eigenvector matrix $\mathbf{Z}_{r_j}$ of $\sum_x$. The estimate of the jth nonlinear pattern $\boldsymbol{f}_j$ is $\mathbf{Z}_{r_j}(\boldsymbol{\Lambda}_{r_j} - \sigma^2\mathbf{I})^{1/2}\boldsymbol{y}^j (j = 1, 2, \ldots, q)$.*

Proof: Equation (15) implies that the $r_j \times 1$ PCA score vector corresponding to the jth nonlinear block is $\boldsymbol{y}^j = (\boldsymbol{\Lambda}_{r_j} - \sigma^2\mathbf{I})^{-1/2}\boldsymbol{s}_j + \boldsymbol{w}_j$, where $\boldsymbol{w}_j = (\boldsymbol{\Lambda}_{r_j} - \sigma^2\mathbf{I})^{-1/2}\mathbf{Z}_{rj}^{\mathrm{T}}\varepsilon$.

Then, $\hat{\boldsymbol{f}}_j$, the estimate of nonlinear pattern $\boldsymbol{f}_j$, can be obtained from $\boldsymbol{y}^j$, because

$$\begin{aligned} \hat{\boldsymbol{f}}_j &= \mathbf{Z}_{r_j}\big(\boldsymbol{\Lambda}_{r_j} - \sigma^2\mathbf{I}\big)^{1/2}\boldsymbol{y}^j = \mathbf{Z}_{r_j}\big(\boldsymbol{\Lambda}_{r_j} - \sigma^2\mathbf{I}\big)^{1/2}\big(\boldsymbol{\Lambda}_{r_j} - \sigma^2\mathbf{I}\big)^{-1/2}\boldsymbol{s}_j + \varepsilon \\ &= \mathbf{E}_j\boldsymbol{s}_j + \varepsilon = \boldsymbol{f}_j + \varepsilon. \end{aligned} \tag{22}$$

The third equation holds in that $\mathbf{E}_j = \mathbf{Z}_{r_j}$ in Eq. (14). □

3.4. *Some implementation issues*

Some issues are considered here to make the proposed nonlinear BSS algorithm performs well for multiple patterns identification. In the case when only sample data is available, the estimate of r is equal to the number of dominant eigenvalues for sample covariance, and the average of the $d - r$ smallest eigenvalues is taken as a robust estimate of σ^2.

As discussed in the theoretical results in Section 3.2, the optimal matrix $\mathbf{V}$ is the minimizer (over all $p \times p$ orthogonal matrices) of the sum of the squares of the entire set of cumulant matrices $\mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}_i)\mathbf{V}$. The advantage of diagonality criterion (7), which is referred to as the joint approximate diagonalization of eigenmatrices (JADE) criterion (Cardoso and Souloumiac, 1994), is that there exists a computationally efficient Jacobi rotation algorithm for finding its minimizer $\mathbf{V}$. Details of the algorithm can be found in Cardoso and Souloumiac (1994). Because of its excellent performance, the JADE algorithm is often used as a benchmark for evaluating other algorithms (Wax and Sheinvald, 1997). Matlab code for the JADE algorithm is available upon request.

The "approximate" term in the acronym comes from the fact that with sample data, no orthogonal transformation will result in the sample joint cumulant matrix exactly the same as in Figure 2. The sample cumulant matrices can only be approximately diagonalized in the sense that Eq. (7) is minimized. The sample cumulants are defined in the obvious way, where the expectations of the quantities are replaced by their sample averages.

4. A sample testing method for clustering diagonal blocks

As shown in Figure 2, all the off-block elements in the optimal joint cumulant matrix $\sum_i \mathbf{V}^{\mathrm{T}}\mathbf{Q}_{\mathrm{Y}}(\mathbf{M}_i)\mathbf{V}$ should be zero. In practice, however, the proposed nonlinear BSS algorithm is applied over sample data, and no matrix $\mathbf{V}$ can produce the ideal result of Figure 2. In other words, the sample cumulant matrices are jointly diagonalized in the sense that criterion (7) is minimized, even though many off-block elements are not equal to zero. The nonzero off-block elements present a problem of correctly identifying the linear and nonlinear blocks from the sample joint cumulant matrix. Accurate clustering of multiple blocks plays a crucial role in linear and nonlinear pattern identification. To improve the separation accuracy and identify each block automatically from the sample result, we propose a statistical testing method for determining the values of p and r_j $(j = 1, 2, \ldots, q)$, as well as clustering the elements of whitened vector $\boldsymbol{y}$ into distinct blocks.

Define the rotated whitened vector as $\boldsymbol{y}^* = \mathbf{V}^{\mathrm{T}}\boldsymbol{y}$. The diagonal structure of the optimal matrix implies that $\boldsymbol{y}^* = [\boldsymbol{v}^{\mathrm{T}}, \boldsymbol{s}_1^{*,\mathrm{T}}, \ldots \boldsymbol{s}_q^{*,\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{s}_j^*$ is a linear transformation of $\boldsymbol{s}_j$ $(j = 1, 2, \ldots, q)$ and is independent of $\boldsymbol{v}$ and $\boldsymbol{s}_i^*$ $(i \neq j)$. In linear BSS, the optimal matrix $\mathbf{V}$ is shown to rotate the elements y_i^* of $\boldsymbol{y}^*$ to be independent $(i = 1, 2, \ldots, r)$ such that all the fourth-order cross-cumulants $\mathrm{Cum}_{i,i,i,j}(y_i^*)$ are zero. This cumulant property is adopted as an independence measure in the proposed block clustering approach: if $\mathrm{Cum}_{i,i,i,j}(\boldsymbol{y}^*) = 0$, then y_i^* and y_j^* are caused by two independent sources; otherwise, they are from a single source t_j and correspond to the same nonlinear block in Figure 2. In this way, the proposed statistical testing method will assign all the elements of $\boldsymbol{y}$ to multiple non-overlapped blocks. The values of p and r_j are determined subsequently.

Define an $r \times r$ sample testing matrix $\mathbf{D}$ on $\boldsymbol{y}^*$ component-wise as

$$
\begin{aligned}
D_{ij} &= \widehat{\mathrm{Cum}}_{i,i,i,j}(\boldsymbol{y}^*) + \widehat{\mathrm{Cum}}_{i,i,j,j}(\boldsymbol{y}^*) + \widehat{\mathrm{Cum}}_{i,j,j,j}(\boldsymbol{y}^*) \equiv D_{i,j}^{(1)} + D_{i,j}^{(2)} + D_{i,j}^{(3)} \\
&= \frac{1}{N}\sum_{n=1}^{N} y_{i,n}^*(y_{j,n}^*)^3 - \frac{3}{N^2}\sum_{n=1}^{N} y_{i,n}^* y_{j,n}^* \sum_{n=1}^{N}(y_{j,n}^*)^2 + \frac{1}{N}\sum_{n=1}^{N}(y_{i,n}^*)^2(y_{j,n}^*)^2
\end{aligned}
$$

$$- \frac{1}{N^2}\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}(y_{j,n}^*)^2 - 2\left(\frac{1}{N}\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right)^2 + \frac{1}{N}\sum_{n=1}^{N}(y_{i,n}^*)^3y_{j,n}^*$$
$$- \frac{3}{N^2}\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\sum_{n=1}^{N}(y_{i,n}^*)^2, \tag{23}$$

where D_{ij} is the ijth entry of $\mathbf{D}$ and symbol "^" denotes the estimate of a quantity. The matrix $\mathbf{D}$ is introduced to investigate the independence of a pair of elements in $\boldsymbol{y}^*$. Under the null hypothesis that y_i^* and y_j^* are independent, the variance of D_{ij} is:

$$\begin{aligned}
\mathrm{Var}(\mathrm{D}_{ij}) &= \mathrm{Var}\big(D_{i,j}^{(1)}\big) + \mathrm{Var}\big(D_{i,j}^{(2)}\big) + \mathrm{Var}\big(D_{i,j}^{(3)}\big) + 2\mathrm{Cov}\big(D_{i,j}^{(1)}, D_{i,j}^{(2)}\big)\\
&\quad + 2\mathrm{Cov}\big(D_{i,j}^{(1)}, D_{i,j}^{(3)}\big) + 2\mathrm{Cov}\big(D_{i,j}^{(2)}, D_{i,j}^{(3)}\big)\\
&= \mathrm{E}\left\{\frac{1}{N^2}\left(\sum_{n=1}^{N}y_{i,n}^*(y_{j,n}^*)^3\right)^2 + \frac{9}{N^4}\left(\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right)^2\left(\sum_{n=1}^{N}(y_{j,n}^*)^2\right)^2\right.\\
&\quad \left. - \frac{6}{N^3}\left(\sum_{n=1}^{N}y_{i,n}^*(y_{j,n}^*)^3\right)^2\sum_{n=1}^{N}(y_{j,n}^*)^2\right\} + E\left\{\frac{1}{N^2}\left(\sum_{n=1}^{N}(y_{i,n}^*)^2(y_{j,n}^*)^2\right)^2\right.\\
&\quad + \frac{1}{N^4}\left(\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}(y_{j,n}^*)^2\right)^2 + \frac{4}{N^4}\left(\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right)^4\\
&\quad - \frac{2}{N^3}\sum_{n=1}^{N}(y_{i,n}^*)^2(y_{j,n}^*)^2\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}(y_{j,n}^*)^2 - \frac{4}{N^3}\sum_{n=1}^{N}(y_{i,n}^*)^2(y_{j,n}^*)^2\\
&\quad \left.\times\left(\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right)^2 + \frac{4}{N^4}\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}(y_{j,n}^*)^2\left(\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right)^2\right\}\\
&\quad + 2\mathrm{E}\left\{\left[\frac{1}{N}\sum_{n=1}^{N}(y_{i,n}^*)^3y_{j,n}^* - \frac{3}{N^2}\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right]\right.\\
&\quad \left.\cdot\left[\frac{1}{N}\sum_{n=1}^{N}(y_{i,n}^*)^2(y_{j,n}^*)^2 - \frac{1}{N^2}\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}(y_{j,n}^*)^2 - \frac{2}{N^2}\left(\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right)^2\right]\right\}\\
&\quad + 2\mathrm{E}\left\{\left[\frac{1}{N}\sum_{n=1}^{N}(y_{i,n}^*)^3y_{j,n}^* - \frac{3}{N^2}\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right]\right.\\
&\quad \left.\cdot\left[\frac{1}{N}\sum_{n=1}^{N}y_{i,n}^*(y_{j,n}^*)^3 - \frac{3}{N^2}\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right]\right\}\\
&\quad + 2\mathrm{E}\left\{\left[\frac{1}{N}\sum_{n=1}^{N}y_{i,n}^*(y_{j,n}^*)^3 - \frac{3}{N^2}\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right]\right.\\
&\quad \left.\cdot\left[\frac{1}{N}\sum_{n=1}^{N}(y_{i,n}^*)^2(y_{j,n}^*)^2 - \frac{1}{N^2}\sum_{n=1}^{N}(y_{i,n}^*)^2\sum_{n=1}^{N}(y_{j,n}^*)^2 - \frac{2}{N^2}\left(\sum_{n=1}^{N}y_{i,n}^*y_{j,n}^*\right)^2\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{N^2-6N+9}{N^3}\mathrm{E}[(y_j^*)^6] + \frac{(N-1)(18-6N)}{N^3}\mathrm{E}[(y_j^*)^4] \\
&\quad + \frac{9(N-1)(N-2)}{N^3} + \frac{N^2-N+4}{N^3}\mathrm{E}[(y_i^*)^4]\mathrm{E}[(y_j^*)^4] \\
&\quad + \frac{(N-1)(5N+12)}{N^3} + \frac{(N-1)(4-N)}{N^3}\{\mathrm{E}[(y_j^*)^4] + \mathrm{E}[(y_i^*)^4]\} \\
&\quad + \frac{N^2-6N+9}{N^3}\mathrm{E}[(y_i^*)^6] + \frac{(N-1)(18-6N)}{N^3}\mathrm{E}[(y_i^*)^4] \\
&\quad + \frac{9(N-1)(N-2)}{N^3} + \frac{2(N^2-3N+6)}{N^3}\mathrm{E}[(y_i^*)^5]\mathrm{E}[(y_j^*)^3] \\
&\quad - \frac{8(N-1)(N-3)}{N^3}\mathrm{E}[(y_i^*)^3]\mathrm{E}[(y_j^*)^3] + \frac{2(9-5N)}{N^3}\mathrm{E}[(y_i^*)^4]\mathrm{E}[(y_j^*)^4] \\
&\quad + \frac{6(N-1)(3-N)}{N^3}\{\mathrm{E}[(y_j^*)^4] + \mathrm{E}[(y_i^*)^4]\} + \frac{12(N-1)(N-2)}{N^3} \\
&\quad + \frac{2(N^2-3N+6)}{N^3}\mathrm{E}[(y_i^*)^3]\mathrm{E}[(y_j^*)^5] \\
&\quad - \frac{8(N-1)(N-3)}{N^3}\mathrm{E}[(y_i^*)^3]\mathrm{E}[(y_j^*)^3]. \qquad (24)
\end{aligned}
$$

Multiplied by the orthogonal matrix $\mathbf{V}$, $\boldsymbol{y}^*$ is still a whitened vector, that is, $\mathrm{E}[(y_i^*)^2] = 1$ in Eq. (24). Under the null hypothesis, when N is large enough D_{ij} is asymptotically normally distributed with zero mean and sample standard deviation $\sqrt{\widehat{\mathrm{Var}}(\mathrm{D}_{ij})}$. The quantity $\widehat{\mathrm{Var}}(\mathrm{D}_{ij})$ is estimated by substituting $\mathrm{E}(\bullet)$ with appropriate sample means. Therefore, for an α-level test (usually $\alpha = 0.05$) of

$\boldsymbol{H}_0$: y_i^* and y_j^* are independent
$\boldsymbol{H}_1$: y_i^* and y_j^* are dependent and from the same nonlinear block,

we reject $\boldsymbol{H}_0$ if $|\mathrm{D}_{ij}| > z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\mathrm{D}_{ij})}$. Here, $z_{\alpha/2}$ is the upper $\alpha/2$ percentage points of a normal distribution. Thus, for any D_{ij} falling outside the confidence interval $[-z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\mathrm{D}_{ij})}, z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\mathrm{D}_{ij})}]$, we have $100(1-\alpha)\%$ confidence that y_i^* and y_j^* are caused by the same source and should be clustered into a single nonlinear block.

Thus far we have presented the results for clustering blocks in Figure 2 and identifying the linear and nonlinear patterns in model (3). The proposed nonlinear BSS algorithm can be summarized as below:

(1) Identify r, the number of dominant eigenvalues of $\boldsymbol{\Sigma}_x$, and set σ^2 equal to the average of the $d - r$ smallest eigenvalues.
(2) Transform $\boldsymbol{x}$ into whitened vector $\boldsymbol{y}$ using Eq. (15).
(3) Calculate the optimal orthogonal matrix $\mathbf{V}$ under criterion (7) using the Jacobi rotation algorithm introduced in Section 3.3.

(4) Define matrix $\mathbf{D}$ for $\boldsymbol{y}^* = \mathbf{V}^{\mathrm{T}}\boldsymbol{y}$ via Eq. (23). Determine p and r_j for the linear and nonlinear blocks in Figure 2 using the hypothesis testing method.
(5) Identify the linear pattern vectors $\boldsymbol{c}_i$ $(i = 1, 2, \ldots, p)$ and recover source $\boldsymbol{v}$ via Eqs. (20) and (21), respectively.
(6) Identify the nonlinear patterns $\boldsymbol{f}_j$ $(j = 1, 2, \ldots, q)$ via Eq. (22).

5. Experimental results

5.1. A synthetic data example

We investigate the performance of proposed nonlinear BSS algorithm on a multivariate data set with varying sample size N and noise variance σ^2 using simulation. The sample data $\boldsymbol{x}$ was generated by the model $\boldsymbol{x} = \boldsymbol{c}_1 v_1 + \boldsymbol{c}_2 v_2 + \boldsymbol{f}_1(t_1) + \boldsymbol{f}_2(t_2) + \varepsilon$, where the independent sources v_1, v_2 are a rectangle and cosine wave.

The nonlinear patterns in this example are assumed to lie in a two-dimensional (i.e., $r_1 = r_2 = 2$) feature space. Therefore, the number of dominant eigenvalues of $\boldsymbol{x}$ is $r = 6$. As shown in Figure 3, the original sources v_1, v_2 were recovered by their esti-

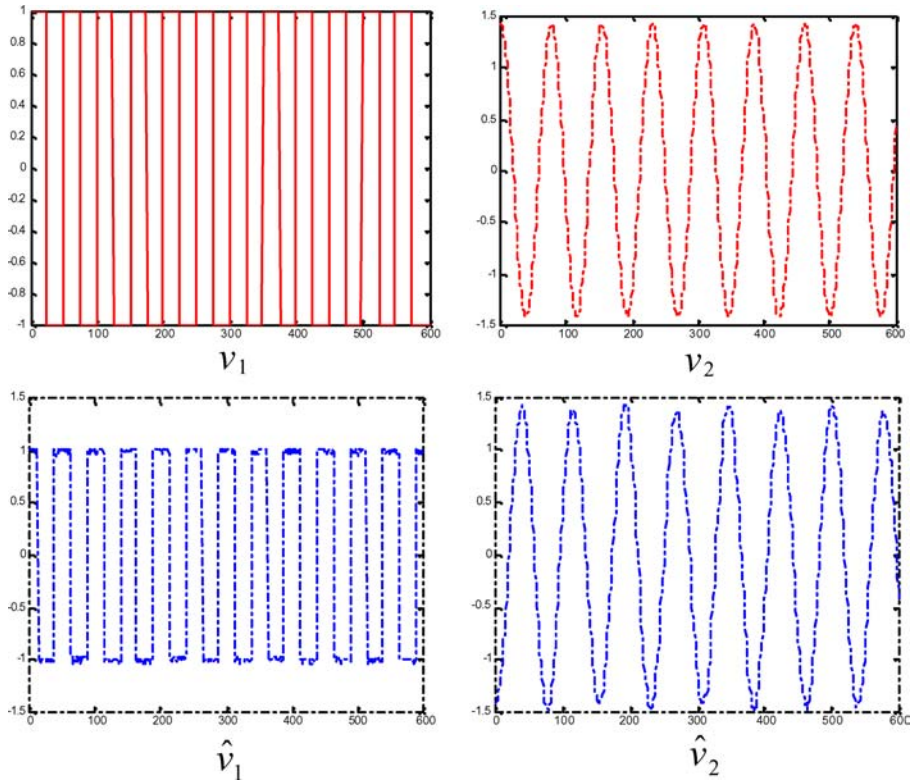

Figure 3. The original and estimated sources by proposed nonlinear BSS algorithm when noise variance $\sigma^2 = 0$.

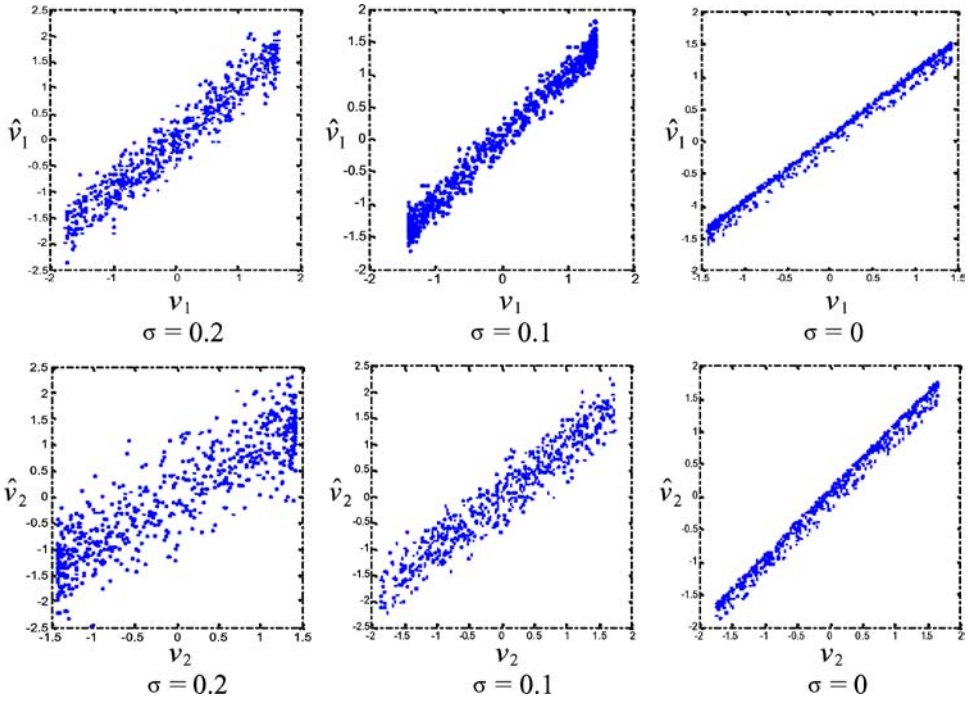

Figure 4. Scatter plots of sources v_1, v_2 versus their estimates with fixed sample size $N = 600$ and varying noise variance σ^2.

mates $\hat{v}_1$ and $\hat{v}_2$. The performance of linear source identification was first illustrated by the scatter plots in Figure 4, which were arranged in the order of increasing noise standard deviations from 0 to 0.2. As expected, the estimation accuracy deteriorated as σ increased.

Quantitative comparison between matrix $\mathbf{C}$ and the estimate $\hat{\mathbf{C}}$ also demonstrates that the proposed nonlinear BSS algorithm is capable of producing accurate linear pattern estimates in multivariate data:

$$\mathbf{C}^{\mathrm{T}} = \begin{bmatrix} -.568 & -.369 & .245 & -.108 & -.203 & .381 & -.119 & -.197 & -.458 \\ -.239 & .406 & -.063 & .235 & .591 & .098 & -.283 & .347 & -.399 \end{bmatrix},$$

$$\hat{\mathbf{C}}^{\mathrm{T}} = \begin{bmatrix} -.542 & -.381 & .228 & -.099 & -.201 & .375 & -.132 & -.197 & -.446 \\ -.221 & .415 & -.072 & .236 & .595 & .091 & -.277 & .355 & -.387 \end{bmatrix}.$$

As discussed in Section 4, the off-block elements in the sample joint cumulant matrix are not necessarily equal to zero. For example, for sample size $N = 600$, $d = 9$ and $\sigma = 0.1$,

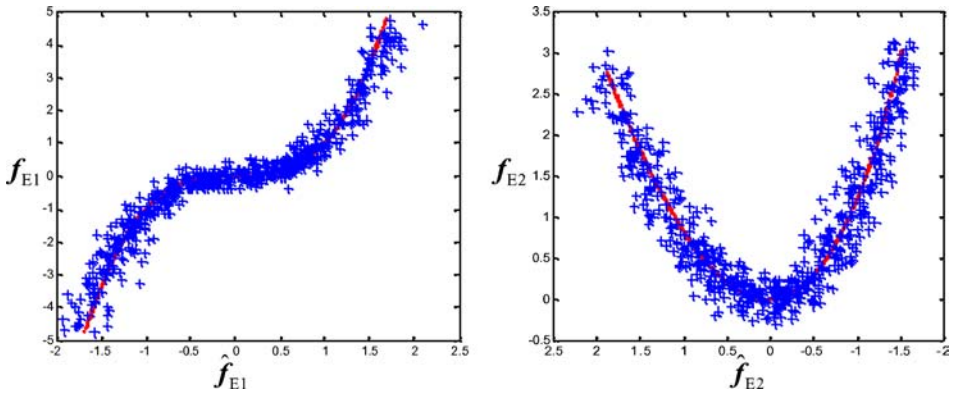

Figure 5. The projection of the original and estimated curves onto the 2D subspace, denoted by "." and "+", respectively. *Left*: f_{E1} and $\hat{f}_{E1}$; *Right*: f_{E2} and $\hat{f}_{E2}$.

the joint cumulant matrix is

$$\begin{bmatrix} .0038 & .0007 & .0016 & .0006 & .0028 & .0008 \\ .0007 & .1434 & .0014 & .0024 & .0007 & .0006 \\ .0016 & .0014 & .1331 & .0669 & .0033 & .0065 \\ .0006 & .0024 & .0669 & .3637 & .0009 & .0028 \\ .0028 & .0007 & .0033 & .0009 & 1.3706 & .379 \\ .0008 & .0006 & .0065 & .0028 & .379 & 9.946 \end{bmatrix}$$

In the block-structured matrix above, it is evident that there are a linear block and two 2×2 nonlinear blocks denoted by dashed lines. The sample testing method proposed in Section 4 verified this observation on block separation.

After clustering the multiple blocks in the sample joint cumulant matrix, the nonlinear patterns were identified using Eq. (22). As assumed in the generative data model, both original curves f_1 and f_2 and their estimates $\hat{f}_1$ and $\hat{f}_2$ lie in a 2D subspace. The projections of these curves onto this feature space are illustrated in Figure 5.

To quantify the estimation accuracy of nonlinear pattern identification, we introduce a Euclidean distance as follows. Let $f_{\mathrm{E}i}$ and $\hat{f}_{\mathrm{E}i}$ be the projection of $f_i(t_i)$ and $\hat{f}_i$ onto the r_i-dimensional subspace ($i = 1, 2$ in this case). Define the squared distance function

$$D_{fi} = \sum_{n=1}^{N} \min_{t_i} \|\hat{f}_{\mathrm{E}i,n} - f_{\mathrm{E}i}(t_i)\|^2, \tag{25}$$

as a performance measure for patterns f_i. Table 1 summarizes the evaluation results with varying d, N, and σ^2, in which the sum of distances measure (25) was averaged over 10,000 Monte Carlo replicates.

Table 1. Performance measure on the closeness between original curve $\boldsymbol{f}_i$ and its estimate $\hat{\boldsymbol{f}}_i$ that were projected onto the 2D subspace ($i = 1, 2$).

d	N	σ	D_{f1}	D_{f2}
6	300	0	0.031	0.046
6	300	0.1	0.056	0.073
6	300	0.2	0.102	0.122
6	600	0	0.049	0.058
6	600	0.1	0.087	0.097
6	600	0.2	0.168	0.186
9	300	0	0.038	0.049
9	300	0.1	0.061	0.081
9	300	0.2	0.117	0.136
9	600	0	0.056	0.075
9	600	0.1	0.091	0.106
9	600	0.2	0.159	0.204
18	300	0	0.042	0.051
18	300	0.1	0.066	0.085
18	300	0.2	0.124	0.141
18	600	0	0.063	0.084
18	600	0.1	0.101	0.118
18	600	0.2	0.169	0.214

As Table 1 shows, the identified curves can characterize the nonlinear features of the original patterns when they lie in a lower-dimensional feature space. Again, estimation accuracy deteriorated with increasing noise variance.

5.2. A real manufacturing example

We now consider the automotive crankshaft manufacturing process, which consists of a number of steps, including forging, rough cutting, finish cutting, drilling, grinding, and polishing. Figure 6 shows the geometry of a crankshaft. In one of the many inspections during manufacturing, stylus traces around the circumference at a number of locations on the main bearings and pin bearings are obtained automatically near the end of production. The difference between the maximum diameter at each location and the target diameter is then logged. The bullet "•" symbols in Figure 6 indicate the locations at which the sensor measurements are taken (Apley & Lee, 2002). The sensors are measured along each of the five main bearings (Mains 1 through 5) and the four pin bearings (Pins 1 through 4).The measurement signal $\boldsymbol{x}$ for each crankshaft consists of $d = 17$ diameter measurements.

Based on a sample of $N = 250$ crankshafts, the proposed nonlinear BSS algorithm revealed that one linear ($p = 1$) and one nonlinear pattern ($q = 1$) are present in $\boldsymbol{x}$.

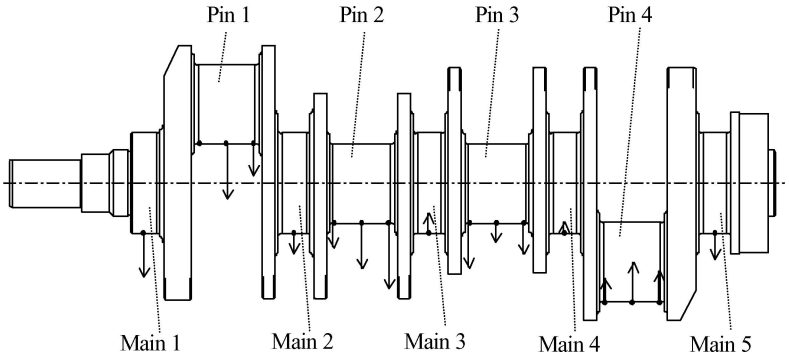

Figure 6. Geometry of a crankshaft with 17 measurement locations, which are denoted by "•".

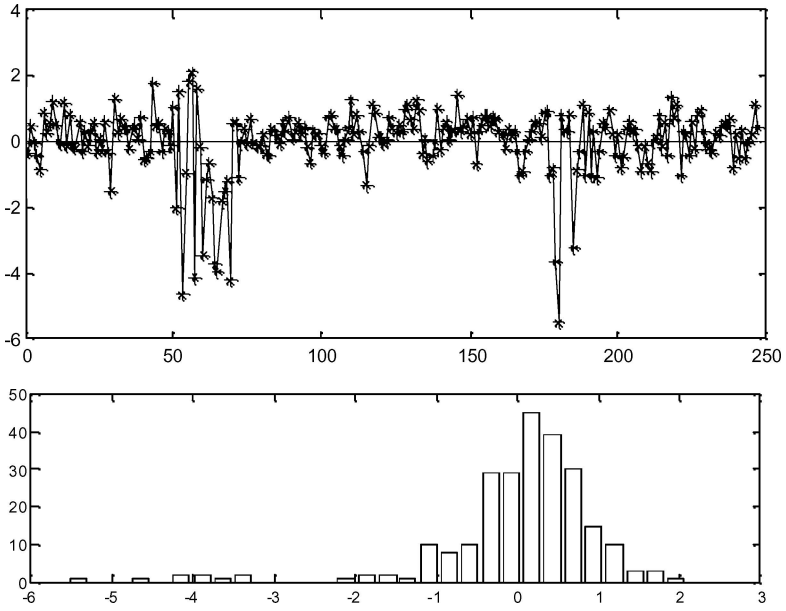

Figure 7. The estimated signal (*top*) and histogram (*bottom*) of source v_1.

Estimates of the linear pattern vector $\boldsymbol{c}_1$ and the corresponding source v_1 are shown in Figures 6 and 7, respectively. Each element of $\hat{\mathbf{c}}_1$ was plotted by a directed arrow at the corresponding measurement point. The length of the arrow is proportional to the magnitude of the element and the direction represents the sign.

The root cause of linear pattern $\boldsymbol{c}_1 v_1$ is some locating elements that failed to constrain the crankshaft properly when it was placed in the subassembly line. The geometry of the part and the position of the locating elements are such that the part is free to deviate by small amounts from its nominal position. The source v_1 is then a random variable that is proportional to

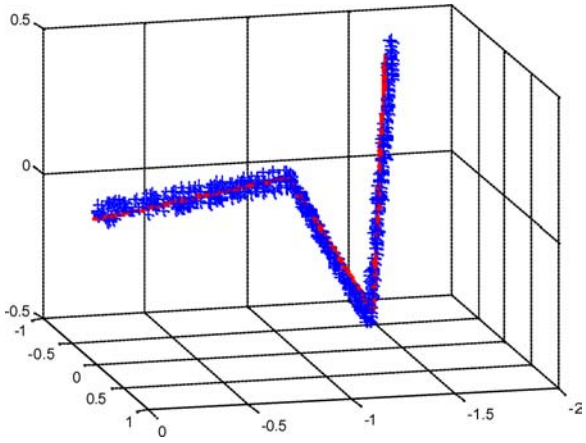

Figure 8. The projection of the identified nonlinear pattern onto a 3D subspace (denoted by "+"), summarized by a principal curve (denoted by ".").

the deviation occurring at the measurement locations. The 17-dimensional vector $\boldsymbol{c}_1$ is determined by the fixture design and measurement deployment, which characterize the spatial nature of the linear effect from source v_1, as shown in Figure 6.

The nonlinear pattern $\boldsymbol{f}_1$ was identified by the proposed algorithm as a piecewise linear curve and its projection onto the 3-D feature space (i.e., $r_1 = 3$) is shown in Figure 8. The presence of nonlinearities in measurement data $\boldsymbol{x}$ can be illustrated by the 4×4 sample joint cumulant matrix as below ($r = r_1 + 1 = 4$):

$$\begin{bmatrix} .095 & .0008 & .0006 & .0012 \\ .0008 & .271 & .075 & .029 \\ .0006 & .075 & .352 & .011 \\ .0012 & .029 & .011 & .768 \end{bmatrix},$$

where the lower right 3×3 submatrix is the nonlinear block corresponding to $\boldsymbol{f}_1$, and the upper left element 0.095 corresponds to $\boldsymbol{c}_1 v_1$. The sample testing method proposed in Section 4 also justified the above clustering result by assigning rotated PCA score elements into a linear and nonlinear block.

The presence of a nonlinear pattern in the measurement data means that it is not proper to apply linear BSS algorithms for pattern analysis, as they may produce erroneous diagnostic information for product quality monitoring and control, as discussed in the introduction. Because the number of dominant eigenvalues of $\boldsymbol{\Sigma}_x$ is $r = 4$, a high-order cumulants-based linear BSS method identified four independent sources contributing to the variability in $\boldsymbol{x}$ (Cardoso, 1998). The first vector $\boldsymbol{c}_1$ found by the linear analysis was the same as that shown in Figure 6; the estimates of the other vectors $\boldsymbol{c}_2$, $\boldsymbol{c}_3$, and $\boldsymbol{c}_4$are shown in Figure 9. As in any process quality control applications, proper identification of statistical patterns and their sources is necessary in order to eliminate the physical root causes, thereby reducing

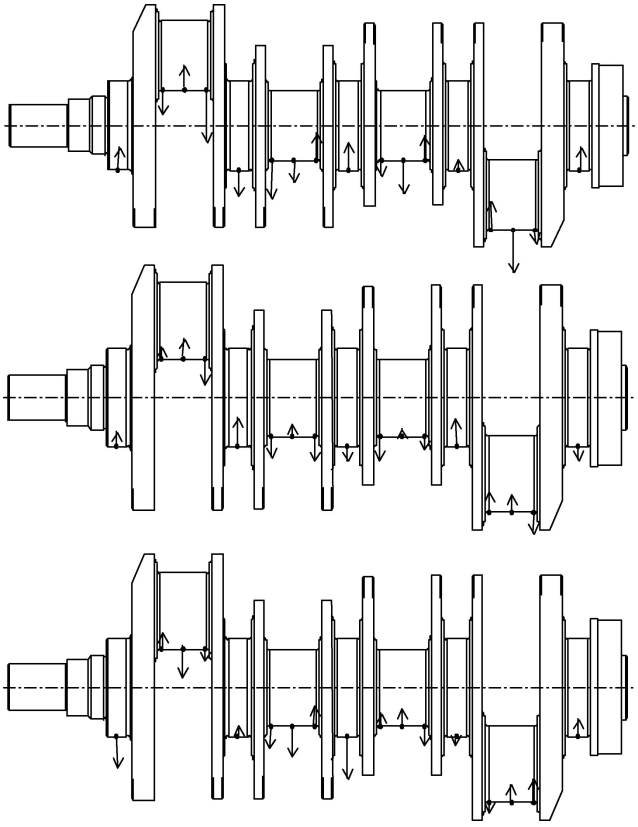

Figure 9. The estimates of $\boldsymbol{c}_2, \boldsymbol{c}_3$, and $\boldsymbol{c}_4$ by the linear BSS method for the crankshaft example are plotted by directed arrows.

process variability. In the present example, the identified $\hat{\boldsymbol{c}}_2$ to $\hat{\boldsymbol{c}}_4$ by linear BSS do not actually correspond to three independent root causes. Unlike the plot of $\hat{\boldsymbol{c}}_1$ in Figure 6, the visualization of these three pattern vectors does not provide intuitive and reasonable explanation of their underlying physical causes. Consequently, the estimated sources $\hat{v}_2$ to $\hat{v}_4$ would present incorrect process diagnosing and monitoring if they were assumed to be single independent sources. On the other hand, because of the complicated physical characteristics, it is not uncommon to have multiple nonlinear patterns in multivariate manufacturing. In this sense, by precisely identifying all the patterns simultaneously present in measurement data, the proposed nonlinear BSS method will serve as a diagnostic aid in facilitating the goal of reducing product variability.

6. Discussion

Although the general problem of blind source separation in nonlinear domains is not solvable, multiple linear and nonlinear pattern separation and identification in a specific and

realistic-enough model, as proposed in this article, is shown to be possible given the same source assumptions as in linear instantaneous BSS. A promising high-order cumulants-based independence criterion is introduced to retrieve the source signals and reveal the nonlinear features embedded in the original data space. When each nonlinear pattern lies in a lower feature space, the joint diagonality criterion is shown to produce a block-structured joint cumulant matrix. The linear and nonlinear patterns are then identified given the optimal rotation matrix. Experimental studies of the performance of the proposed nonlinear BSS algorithm, including a real example from crankshaft manufacturing, demonstrate its potential use in multivariate data analysis.

In the proposed nonlinear model, $\boldsymbol{f}_j$ are assumed to be one-dimensional curves (i.e., principal curves). This applies to the situations when each nonlinear pattern is a function of single variation source t_j, and may lead to erroneous identification results when $\boldsymbol{f}_j$ are higher-dimensional manifolds (e.g., $\boldsymbol{f}_j$ is a principal surface of multiple sources). Such a more generic model requires the estimation of the intrinsic dimension of the manifolds, as well as the visualization of identified patterns, which is a challenging problem for future research.

The proposed nonlinear source separation method will be extended in several ways in the future. An obvious extension is the inclusion of time delays into the nonlinear BSS model, which will take advantage of temporal information often present in the source signals and is expected to help in describing the nonlinear patterns more precisely.

Another possible development involves the extension of nonlinear models with locally varying measurement noise, i.e., where the parameters of the noise covariance matrix may depend on different spatial measurement features. This is desirable because using a fixed isotropic noise covariance does not fit the data well in some applications.

References

Apley, D. W., & Lee, H. Y. (2002). Identifying spatial variation patterns in multivariate manufacturing processes: A blind separation approach. *Technometrics* (submitted).

Burel, G. (1992). Blind Separation of sources: A nonlinear neural algorithm. *Neural Networks, 5*, 937–947.

Cardoso, J. F. (1998). Blind signal separation: Statistical principles. *Proceedings of the IEEE, 86*, 2009–2025.

Cardoso, J. F., & Souloumiac, A. (1994). Blind beamforming for non-gaussian signals. *IEE Proceedings, 140*, 362–370.

Chang, K., & Ghosh, J. (1998). Principal curves for nonlinear feature extraction and classification. *SPIE Applications of Artificial Neural Networks in Image Processing III, 3307*, 120–129.

Deco, G., & Brauer, W. (1995). Nonlinear higher-order statistical decorrelation by volume-conserving architectures. *Neural Networks, 8*, 525–535.

Hastie, T., & Stuetzle, W. (1989). Principal Curves. *Journal of the American Statistical Association, 84*, 502–516.

Hyvarinen, A., & Pajunen, P. (1999). Nonlinear independent component analysis: Existence and uniqueness results. *Neural Networks, 11*, 429–439.

Hyvarinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*, NY:Wiley.

Jutten, C., & Karhunen, J. (2003). Advances in nonlinear blind source separation. In *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Source Separation*, Nara, Japan (pp. 245–256).

Krob, M., & Benidir, M. (1994). Separation of a polynomial mixture of independent sources using higher-order output moment matrices. In *Proc. Signal Process. VII: Theories Applicat.*, vol. 1 (pp. 187–190).

Lee, T. W., Koehler, B., & Orglmeister, R. (1997). Blind source separation of nonlinear mixing models. In *Neural Networks for Signal Processing VII*. IEEE Press (pp. 406–415).

Pajunen, P., Hyvarinen, A., & Karhunen, J. (1996). Nonlinear source separation by self-organizing maps. In *Proc. ICONIP*, vol. 2, Hong Kong, Sept. 1996 (pp. 1207–1210).

Pajunen, P., & Karhunen, J. (1997). A maximum likelihood approach to nonlinear blind source separation. In *Proc. ICANN*, Lausanne, Switzerland (pp. 541–546).

Taleb, A. (2002). A generic framework for blind source separation in structured nonlinear models. *IEEE Trans. Signal Processing, 50:8*, 1819–1830.

Taleb, A., & Jutten, C. (1999). Source separation in post nonlinear mixtures. *IEEE Trans. Signal Processing, 47* 2807–2820.

Valpola, H., et al. (2003). Nonlinear blind source separation by variational bayesian learning. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *E86:3*, 532–541.

Valpola, H., & Karhunen, J. (2002). An unsupervised ensemble learning method for nonlinear dynamic state-space models. *Neural Computation*, *14:11*, 1647–1692.

Wax, M., & Sheinvald, J. (1997). A least-squares approach to joint diagonalization. *IEEE Signal Processing Letters, 4*, 52–53.

Yang, H. H., Amari, S., & Cichocki, A. (1998). Information-theoretic approach to blind separation of sources in nonlinear mixture. *Signal Processing, 64:3*, 291–300.

Received July 22, 2003
Revised January 31, 2005
Accepted April 3, 2005