IIE Transactions
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uiie20

Separation of individual operation signals from mixed sensor measurements

Qingyu Yang [a] & Jionghua (Judy) Jin [b]

[a] Department of Industrial and System Engineering , Wayne State University , Detroit , MI , 48202 , USA

[b] Department of Industrial and Operations Engineering , The University of Michigan , Ann Arbor , MI , 48109 , USA

Accepted author version posted online: 08 Aug 2011.Published online: 14 Jun 2012.

To cite this article: Qingyu Yang & Jionghua (Judy) Jin (2012) Separation of individual operation signals from mixed sensor measurements, IIE Transactions, 44:9, 780-792, DOI: 10.1080/0740817X.2011.609873

To link to this article: http://dx.doi.org/10.1080/0740817X.2011.609873

IIE Transactions (2012) **44**, 780–792
Copyright © "IIE"
ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2011.609873

Separation of individual operation signals from mixed sensor measurements

QINGYU YANG[1] and JIONGHUA (JUDY) JIN[2*]

[1]*Department of Industrial and System Engineering, Wayne State University, Detroit, MI 48202, USA*
[2]*Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109, USA*
E-mail: jhjin@umich.edu

Received March 2010 and accepted June 2011

Sensor system measurements are generally mixed signals measured from multiple independent/dependent operations embedded in a complex system. In this article, a novel method is developed to separate the source signals of individual operations from the mixed sensor measurements by integrating the independent component analysis method and the Sparse Component Analysis (SCA) method. The proposed method can efficiently estimate the source signals that include both independent signals and dependent signals that have some dominant components in the time or some linear transform domains (e.g., frequency domain, time/frequency domain, or wavelet domain). In addition, an SCA method is also developed in this article that can automatically identify the dominant components in multiple linear transform domains. A case study of a forging process is conducted to demonstrate the effectiveness of the proposed methods.

Keywords: Blind source separation, independent component analysis, mixed sensor signal, sparse component analysis

1. Introduction

In general, complex manufacturing processes consist of multiple independent/dependent operations, and it quickly becomes impossible or simply unaffordable to directly measure all individual operations separately. The available sensor measurements are usually the combined responses of multiple operations. In this article, the responses of individual operations are called *source signals*, and the sensor measurements are called *mixed sensor signals*. If there is a linear relationship between the source signals $\mathbf{S}(t)$ and the mixed sensor signals $\mathbf{X}(t)$, the following equation holds:

$$\mathbf{X}(t) = \mathbf{A}\mathbf{S}(t) + \varepsilon(t), \quad (1)$$

where t denotes the time index within a complete cycle of repetitive system operations. A discrete sampling time t is used for both mixed sensor signals and source signals; i.e., $t = 1, 2, \ldots, N$, where N is the number of data points sampled within a complete cycle of system operations. $\mathbf{S}(t) \equiv [S_1(t)\ S_2(t), \ldots, S_n(t)]^{\mathrm{T}} \in \Re^{n\times N}$ represents n unknown source signals corresponding to n individual operations embedded in a system; $\mathbf{X}(t) \equiv [X_1(t)\ X_2(t), \ldots, X_m(t)]^{\mathrm{T}} \in \Re^{m\times N}$represents m $(m \geq n)$ mixed sensor signals; $\mathbf{A} \in \Re^{m\times n}$is an unknown constant mixing matrix, representing the linear relationship between the mixed sensor signals and the source signals; and $\varepsilon(t) \equiv [\varepsilon_1(t)\ \varepsilon_2(t), \ldots, \varepsilon_m(t)]^{\mathrm{T}} \in \Re^{m\times N}$ represents m sensor noises following a multivariate independent and identically distributed normal distribution; i.e., $\varepsilon(t) \sim MN(0, \zeta^2\mathbf{I})$, where $\mathbf{I}$ is an $m \times m$ identity matrix.

The mixed sensor signals, which are the combined responses of the source signals, are generally shown as a high-dimensional functional data structure and have complex non-stationary characteristics in both the time and frequency domains. Although recently developed advanced Statistical Process Control (SPC) methods can be successfully used to detect system abnormalities based on mixed sensor signals, it is still difficult and/or time consuming to find out which embedded operation has created the system abnormality. In contrast, if we can separate individual source signals from the mixed sensor signals, it will become feasible to directly monitor each source signal and provide explicit diagnostic information for individual operations. Therefore, there is considerable interest in research on estimating source signals from mixed sensor signals.

For the purpose of estimating individual source signals from the mixed sensor signals when the constant mixing matrix $\mathbf{A}$ in Equation (1) is unknown, Independent Component Analysis (ICA; Herault and Jutten, 1986; Gaeta and Lacoume, 1990; Comon, 1994; Hyvarinen and Oja, 1997; Hyvrinen *et al.*, 2001) and Sparse Component Analysis (SCA; Jourjine *et al.*, 2000; Lewicki and Sejnowski, 2000; Bofill and Zibulevsky, 2001; Zibulevsky

*Corresponding author

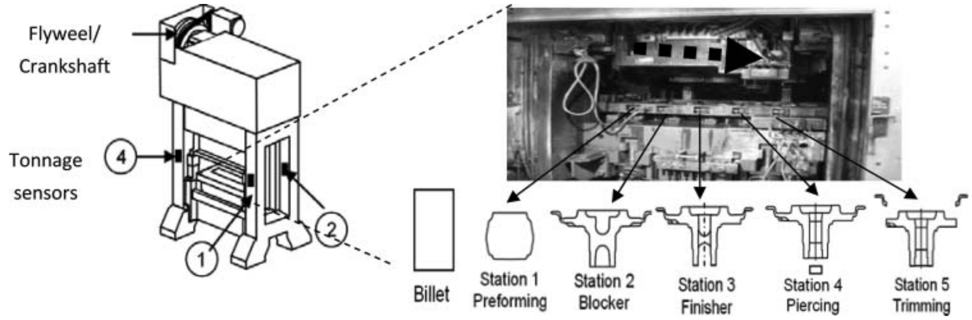

Fig. 1. A multiple operation forging process consisting of five operations.

and Pearlmutter, 2001; Yilmaz and Rickard, 2004; Abrard and Deville, 2005; Li *et al.*, 2006) are two general methods studied in the literature.

The ICA method is used to effectively separate individual source signals from mixed sensor signals based on the assumption that the source signals are statistically independent and non-Gaussian. The ICA method was first proposed by Herault and Jutten (1986) and used to separate independent source signals from mixed sensor signals using a recurrent artificial neural network. Later, Comon (1994) found that mutual information was a suitable measure of "independence," and he further developed an ICA algorithm by minimizing the mutual information between the estimated source signals. Furthermore, Hyvarinen and Oja (1997) developed a fast ICA method by maximizing the non-Gaussian nature of the estimated source signals, which was shown to be equivalent to minimizing the mutual information between source signals. Gaeta and Lacoume (1990) proposed an ICA method based on maximum likelihood estimation that was shown to be essentially equivalent to the ICA method based on mutual information and a non-Gaussian nature.

In contrast, the SCA method is often used to separate source signals from mixed sensor signals no matter whether or not the source signals are dependent. The source signals are called sparse signals if only one source signal occurs (or is dominant) at either specific time periods or feature subsets in a linear transform domain (i.e., a single source signal provides a sole or a dominant contribution to the mixed sensor signals at either these time periods or feature subsets). The first SCA method was proposed by Lewicki and Sejnowski (2000). Jourjine *et al.* (2000) proposed the DUET algorithm, which was based on a Short-Time Frequency Transform (STFT), which assumed that all of the feature subsets of the mixed sensor signals in the time/frequency domain were contributed to by only one source signal. Abrard and Deville (2005) improved this method by using a less restrictive assumption; i.e., for each source signal, there exists a feature subset in the time/frequency region for which only this source signal occurs. Recently, Li *et al.* (2006) proposed a method to estimate source signals in the wavelet domain.

Although significant successes have been achieved in both the ICA and SCA fields, the application of ICA and SCA methods in practice is still limited. The major obstacle is that a complex system usually cannot completely satisfy the assumption that all source signals are independent of each other or each of the individual source signals is dominant at either specific time periods or feature subsets in a linear transform domain.

For example, a progressive forging process, as shown in Fig. 1, consists of five embedded dies that perform different operations at five separate stations: preforming, blocker, finisher, piercing, and trimming. An automatic part-feeding device is used to load the raw material billet and advance the intermediate part between processing stations. All five operations are performed within each press stroke through the five embedded dies exerting tonnage forces on the billet or intermediate parts at each corresponding station. A final product is produced after each billet has sequentially passed through all five stations.

For process monitoring purposes, strain gage sensors are usually installed on the columns of a forging press machine to measure the press tonnage forces. The force measurements from the gage sensors can be considered to be mixed sensor signals that are linear combinations of the die forces (i.e., source signals) generated at the individual operations. Figure 2 shows the individual die forces generated by these five die operations. It has been reported in the forging/stamping processes literature that the active working range of each operation is only a limited portion of the complete operation cycle (Jin, 2004; Jin and Shi, 2005). For the forging process illustrated in Fig. 1, the active working range of the ith individual operation is denoted by $\mathbf{L}_i \equiv [l_{i,0}, l_{i,1}]$, $i = 1, 2, \ldots, 5$, and the values of $\mathbf{L}_i$ are listed in Table 1, in which $l_{i,0}$ and $l_{i,1}$ ($i = 1, 2, \ldots, 5$) denote the start and end time indices of the active working range of station i. Furthermore, data segmentation is conducted to divide the complete operation cycle into 11 data segments; i.e., $\mathbf{T}_i$, $i = 1, 2, \ldots, 11$, with each boundary of a segment being either a start or an end time index of the active working ranges (Jin, 2004). This data segmentation method allows the stations contributing to a given data segment to be explicitly known. For example, segment $\mathbf{T}_2$

Table 1. Active working ranges of the five operations embedded in the forging process illustrated in Fig. 1

L_1	L_2	L_3	L_4	L_5
[35, 494]	[179,672]	[436,720]	[196,353]	[100,268]

is only contributed to by station 1, whereas segment $\mathbf{T}_3$ is contributed to by both stations 1 and 5.

Furthermore, based on die design knowledge, it is known that the die forces generated by the piercing operation and the trimming operation are two independent source signals, whereas the die forces corresponding to the preforming, blocker, and finisher operations are three dependent source signals. If there is no engineering knowledge available, statistical methods available in the literature (Chiu *et al.*, 2003; Karvanen, 2005) can be used to offline check whether source signals are independent or dependent.

Because the source signals consist of not only independent source signals but also dependent source signals, the assumption of the standard ICA method, which assumes that all of the source signals are independent, is not fully satisfied. Also, it can be seen from Fig. 2 that the magnitudes of the independent source signals generated at stations 4 and 5 are much smaller than those of the dependent source signals generated at stations 1 to 3. This further complicates the problem of finding the specific time periods or feature subsets in a linear transform domain in which the source signals generated at stations 4 and 5 dominantly contribute to the mixed sensor signals. Therefore, the assumption that underpins the SCA method cannot be fully satisfied. As a result, neither a single ICA method nor an SCA method can be used to fully separate these five individual die forces from mixed tonnage sensor measurements.

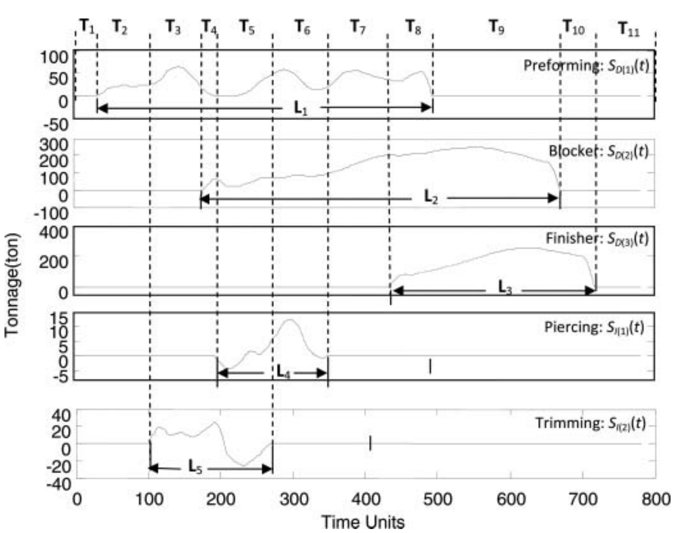

Fig. 2. Source signals generated for individual operations.

A novel signal separation method is developed in this article for source signals that include both independent and dependent signals. For the dependent source signals, it is assumed that there are some dominant components in some linear transform domains. In the proposed method, the integration of ICA and SCA is achieved through two major steps: in step 1, the ICA method is applied to the mixed sensor signals; the estimated independent source signals are then identified from the estimated signals after using the ICA method. In step 2, the impact of the independent source signals is removed from the mixed sensor signals, and the remaining sensor signals, called the reduced sensor signals, are used to further estimate the dependent source signals using the SCA method.

The rest of this article is organized as follows: Section 2 introduces an overview of the proposed methods. The details of the methodology development are presented in Sections 3 and 4: Section 3 discusses the ICA method for estimating the independent source signals from the linear mixed sensor signals and Section 4 develops a new SCA method to estimate the dependent source signals. A case study is investigated in Section 5, and conclusions are drawn in Section 6.

2. Overview of the proposed methods

2.1. *Mathematical model of the signals and assumptions*

When considering both the independent source signals and the dependent source signals, model (1) can be rewritten as follows:

$$\mathbf{X}(t) = [\mathbf{A}_{\mathrm{I}}\mathbf{A}_{\mathrm{D}}] \times \begin{bmatrix} \mathbf{S}_{\mathrm{I}}(t) \\ \mathbf{S}_{\mathrm{D}}(t) \end{bmatrix} + \varepsilon(t), \tag{2}$$

where $\mathbf{S}_{\mathrm{I}}(t) \equiv [S_{\mathrm{I}(1)}(t) S_{\mathrm{I}(2)}(t) \cdots S_{\mathrm{I}(p)}(t)]^{\mathrm{T}} \in \Re^{p\times N}$ represents p independent source signals that are statistically independent of other source signals. $\mathbf{S}_{\mathrm{D}}(t) \equiv [S_{\mathrm{D}(1)}(t) S_{\mathrm{D}(2)}(t) \cdots S_{\mathrm{D}(q)}(t)]^{\mathrm{T}} \in \Re^{q\times N}$, $p + q = n$, represents q dependent source signals that are statistically dependent because these corresponding operations may share some common physical factors. In this article, the numbers p and q are assumed to be given. In addition, both the independent source signals $\mathbf{S}_{\mathrm{I}}(t)$ and the dependent source signals $\mathbf{S}_{\mathrm{D}}(t)$ are assumed to be non-Gaussian; $\mathbf{A}_{\mathrm{I}} \equiv [\mathbf{a}_{\mathrm{I}(1)}\mathbf{a}_{\mathrm{I}(2)} \cdots \mathbf{a}_{\mathrm{I}(p)}] \in \Re^{m\times p}$ and $\mathbf{A}_{\mathrm{D}} \equiv [\mathbf{a}_{\mathrm{D}(1)}\mathbf{a}_{\mathrm{D}(2)} \cdots \mathbf{a}_{\mathrm{D}(q)}] \in \Re^{m\times q}$ are unknown constant mixing matrices, representing the linear relationship between the mixed sensor signals and the independent/dependent source signals, respectively. In order to fully estimate the source signals, both $\mathbf{A}_{\mathrm{I}}$ and $\mathbf{A}_{\mathrm{D}}$ are assumed to have full column rank. For convenience, and without loss of generalization, columns $\mathbf{A}_{\mathrm{I}}$ and $\mathbf{A}_{\mathrm{D}}$ are scaled so that $(\mathbf{a}_{\mathrm{I}(j)})^{\mathrm{T}}\mathbf{a}_{\mathrm{I}(j)} = 1$ and $(\mathbf{a}_{\mathrm{D}(i)})^{\mathrm{T}}\mathbf{a}_{\mathrm{D}(i)} = 1$. $\mathbf{X}(t)$ and $\varepsilon(t)$ are defined as before, and $\varepsilon(t)$ is assumed to be statistically independent of both $\mathbf{S}_{\mathrm{I}}(t)$ and $\mathbf{S}_{\mathrm{D}}(t)$.

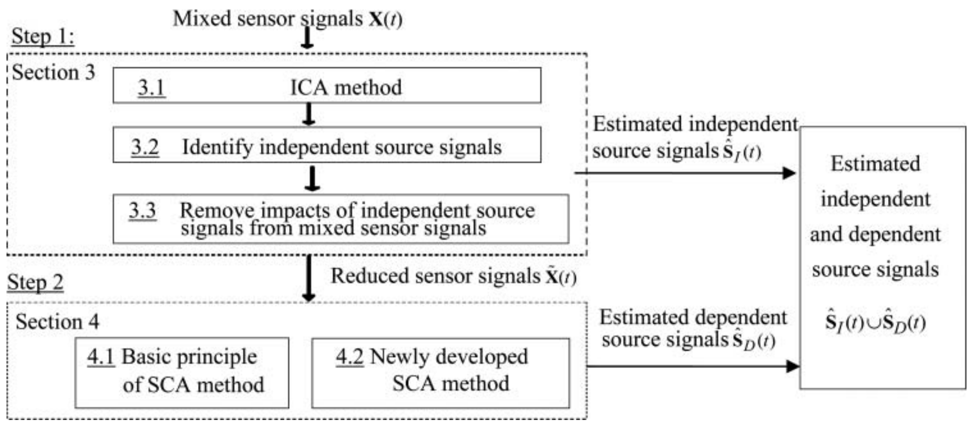

Fig. 3. Method for estimating the independent/dependent source signals from mixed sensor signals.

2.2. *Proposed two-step analysis framework*

There are two major steps, as shown in Fig. 3, in the development of the method to estimate individual independent/dependent source signals from mixed sensor signals. In step 1 to be discussed in Section 3, the estimated independent source signals, denoted as $\hat{\mathbf{S}}_{\mathrm{I}}(t)$, are obtained based on the ICA method. Specifically, in Section 3.1, the ICA method is applied to the mixed sensor signals $\mathbf{X}(t)$. Since the assumption of the ICA method is not satisfied for all source signals, it will be proved that the estimated signals after using the ICA method, denoted as $\mathbf{Y}(t)$, will surely include the estimated independent source signals $\hat{\mathbf{S}}_{\mathrm{I}}(t)$. In Section 3.2, an algorithm is developed to identify $\hat{\mathbf{S}}_{\mathrm{I}}(t)$ from $\mathbf{Y}(t)$ that estimates the statistically independent source signals by maximizing the "independence" of the estimated signals. Then, in Section 3.3, the impact of the independent source signals is removed from the mixed sensor signals. In step 2 to be discussed in Section 4, the remaining sensing signals, called the reduced sensor signals and denoted as $\tilde{\mathbf{X}}(t) \equiv [\tilde{X}_1(t), \tilde{X}_2(t), \ldots, \tilde{X}_m(t)]^{\mathrm{T}} \in \Re^{m\times N}$, are used to obtain the estimated dependent source signals, denoted as $\hat{\mathbf{S}}_{\mathrm{D}}(t)$, by using the SCA method. Specifically, Section 4.1 introduces the SCA method and a novel SCA procedure is developed in Section 4.2. After these two steps of separation analyses, both independent and dependent source signals are separately estimated from the mixed sensor signals.

3. Estimation of independent source signals using ICA

The main idea of the ICA method is to find a constant unmixing matrix $\mathbf{W} \in \Re^{n\times m}$such that

$$\mathbf{Y}(t) = \mathbf{W}\mathbf{X}(t), \tag{3}$$

where $\mathbf{Y}(t) \equiv [Y_1(t), Y_2(t), \ldots, Y_n(t)] \in \Re^{n\times N}$ estimates the statistically independent source signals by maximizing the independence of the estimated signals. In the ICA literature (Hyvrinen *et al.*, 2001), the independence is quantitatively measured by using *kurtosis*, *negentropy*, or *entropy*.

3.1. *ICA method applied to the mixed sensor signals*

In this research, the source signals are composed of both independent source signals $\mathbf{S}_{\mathrm{I}}(t)$ and dependent source signals $\mathbf{S}_{\mathrm{D}}(t)$. Although the assumption of the ICA method is not satisfied, the ICA method can still be applied to the mixed sensor signals $\mathbf{X}(t)$ based on the following Proposition 1, which shows that the estimated signals after using the ICA method include the estimate of the independent source signals; i.e., $\hat{\mathbf{S}}_{\mathrm{I}}(t)$. Please refer to the Appendix for a detailed proof of Proposition 1.

Proposition 1. *The estimated signals after applying the ICA method on the linear mixed sensor signals* $\mathbf{X}(t)$ *are composed of* (i) *a set of p signals* $\hat{\mathbf{S}}_{\mathrm{I}}(t)$ *that estimate independent source signals* $\mathbf{S}_{\mathrm{I}}(t)$ *and* (ii) *a set of q signals, denoted by* $\hat{\mathbf{S}}_{\mathrm{D'}}(t)$. $\hat{\mathbf{S}}_{\mathrm{D'}}(t) \in \Re^{q\times N}$ *estimates signals that are a linear transformation of* $\mathbf{S}_{\mathrm{D}}(t)$; *i.e.,* $\mathbf{D}\mathbf{S}_{\mathrm{D}}(t)$, *where* $\boldsymbol{D} \in \Re^{q\times q}$ *is an unknown constant matrix that maximizes the independence of* $\mathbf{D}\mathbf{S}_{\mathrm{D}}(t)$.

3.2. *Identification of estimated independent source signals*

Based on Proposition 1, the estimated signals after using the ICA method—i.e., $\mathbf{Y}(t)$—include not only $\hat{\mathbf{S}}_{\mathrm{I}}(t)$ but also $\hat{\mathbf{S}}_{\mathrm{D'}}(t)$. In this subsection, an algorithm is developed to identify $\hat{\mathbf{S}}_{\mathrm{I}}(t)$ from $\mathbf{Y}(t)$ based on the active working ranges of different operations. In this article, it is assumed that the active working ranges of all independent/dependent source signals do not exactly overlap and are pre-known from engineering design knowledge. Since the magnitude of source signals is zero outside the active working ranges, if $Y_i(t)$ estimates the jth independent source signal $S_{\mathrm{I}(j)}(t)$, the energy of $Y_i(t)$ should concentrate in the active working range of $S_{\mathrm{I}(j)}(t)$, denoted as $\mathbf{L}_{\mathrm{I}(j)} \equiv [l_{\mathrm{I}(j),0}, l_{\mathrm{I}(j),1}], j = 1, 2, \ldots,$

p. Here, the energy of signal $Y_i(t)$ is defined as

$$E_{i,0} = \sum_{t=1}^{N} (Y_i(t))^2, \tag{4}$$

and the energy of signal $Y_i(t)$ in segment $\mathbf{L}_{\mathrm{I}(j)}$, denoted as $E_{i,j}$, is calculated as follows:

$$E_{i,j} = \sum_{t=l_{\mathrm{I}(j),0}}^{l_{\mathrm{I}(j),1}} (Y_i(t))^2, \tag{5}$$

Define $\eta_{i,j}$ as the ratio of $E_{i,j}$ to $E_{i,0}$, $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, p$; that is,

$$\eta_{i,j} = \frac{E_{i,j}}{E_{i,0}}, \tag{6}$$

and $\eta_{i,j}$ represents the degree of matching between $Y_i(t)$ and $S_{\mathrm{I}(j)}(t)$. In addition, $\eta_{i,j} \in [0, 1]$ and a larger value of $\eta_{i,j}$ indicate a better matching. Based on $\eta_{i,j}$, the detailed steps to identify $\hat{\mathbf{S}}_{\mathrm{I}}(t)$ from $\mathbf{Y}(t)$ can be listed as follows:

Step 1. Initially set the set of estimated independent source signals $\hat{\mathbf{S}}_{\mathrm{I}}(t) = \emptyset$, and the set of remaining signals $\Gamma = \{Y_1(t), Y_2(t), \ldots, Y_n(t)\}$.
Step 2. For each active working range of individual independent operations—i.e., $\mathbf{L}_{\mathrm{I}(j)} \equiv [l_{\mathrm{I}(j),0}, l_{\mathrm{I}(j),1}]$, $j = 1, 2, \ldots, p$—do Steps 3 to 5.
Step 3. For each $Y_i(t)$ in Γ, calculate $\eta_{i,j}$ based on Equation (6).
Step 4. Find the largest $\eta_{i,j}$, denoted as $\eta_{k,j}$; thus, $Y_k(t)$ estimates the jth independent source signal; i.e., $S_{\mathrm{I}(j)}(t)$.
Step 5. Set $\hat{\mathbf{S}}_{\mathrm{I}}(t) = \hat{\mathbf{S}}_{\mathrm{I}}(t) \cup Y_k(t)$ and $\Gamma = \Gamma - Y_k(t)$.

3.3. *Remove independent source signals from mixed sensor signals*

Although the linear transformation of dependent source signals—i.e., $\mathbf{DS}_{\mathrm{D}}(t)$, is estimated after applying the ICA method on the mixed sensor signals, the actual dependent source signals $\mathbf{S}_{\mathrm{D}}(t)$ cannot be directly obtained because the constant matrix $\mathbf{D}$ is unknown. To estimate $\mathbf{S}_{\mathrm{D}}(t)$, the impact of the independent source signals should be eliminated from the mixed sensor signals. The resultant reduced sensor signals after removing the impact of the independent source signals are called the reduced sensor signals denoted as $\tilde{\mathbf{X}}(t)$. $\tilde{\mathbf{X}}(t)$ is subsequently used to estimate the dependent source signals $\hat{\mathbf{S}}_{\mathrm{D}}(t)$ based on the SCA method.

The following Corollary 1 shows the method to obtain $\tilde{\mathbf{X}}(t)$ and the detailed proof of Corollary 1 is given in the Appendix.

Corollary 1. *The reduced sensor signals—i.e.,* $\tilde{\mathbf{X}}(t)$*,* $t = 1, 2, \ldots, N$*—can be calculated as follows:*

$$\tilde{\mathbf{X}}(t) = \mathbf{U}_{\mathrm{D}}\hat{\mathbf{S}}_{\mathrm{D}'}(t) = \mathbf{X}(t) - \mathbf{U}_{\mathrm{I}}\hat{\mathbf{S}}_{\mathrm{I}}(t), \tag{7}$$

where $\mathbf{U}_{\mathrm{D}} \in \Re^{m\times q}$ *is a matrix that is composed of the column vectors of* $\mathbf{U}$ *corresponding to the signals in* $\hat{\mathbf{S}}_{\mathrm{D}'}(t)$*, and* $\mathbf{U}$ *is calculated as* $\mathbf{U} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}} \in \Re^{m\times n}$*;* $\mathbf{U}_{\mathrm{I}} \in \Re^{m\times p}$ *is a matrix composed of the column vectors of* $\mathbf{U}$ *that correspond to the source signals in* $\hat{\mathbf{S}}_{\mathrm{I}}(t)$*;* $\boldsymbol{W}$ *is the unmixing matrix in Equation* (3)*, which is obtained after applying the ICA method on the linear mixed sensor signals* $\mathbf{X}(t)$*.*

Due to sensor noises and the estimation errors of the ICA method, each of the reduced sensor signals $\tilde{\mathbf{X}}(t)$ obtained from Corollary 1—i.e., $\tilde{X}_i(t)$, $i = 1, 2, \ldots, m$—can be modeled as follows:

$$\tilde{X}_i(t) = \sum_{k=1}^{q} A_{i,\mathrm{D}(k)} S_{\mathrm{D}_k}(t) + \tau_i(t) \tag{8}$$

where $A_{i,\mathrm{D}(k)}$ is the element of $\mathbf{A}_{\mathrm{D}}$ at the ith row and the kth column, $i = 1, 2, \ldots, m$, and $k = 1, 2, \ldots, q$; and $\tau_i(t)$, $i = 1, 2, \ldots, m$, indicates the estimation error.

4. Estimation of dependent source signals using SCA

Based on the reduced sensor signals $\tilde{\mathbf{X}}(t)$, the estimation of the dependent source signals—i.e., $\hat{\mathbf{S}}_{\mathrm{D}}(t)$—will be further obtained by using the SCA method in the time domain or in some linear transform domains. The basic idea of the SCA method is to first estimate $\mathbf{A}_{\mathrm{D}}$; i.e., the kth column of $\mathbf{A}_{\mathrm{D}}$ is estimable if the kth dependent source signal $S_{\mathrm{D}(k)}(t)$ is known to be the dominant signal at a given time period or a feature subset in a linear transform domain. After obtaining $\mathbf{A}_{\mathrm{D}}$, the dependent source signals can be further estimated by using a statistical method; e.g., the least squares estimation method based on Equation (8).

4.1. *Basic principle of SCA in the time or a linear transform domain*

Suppose that the kth dependent source signal $S_{\mathrm{D}(k)}(t)$ is dominant at a given time period $[t_{k,a}\ t_{k,b}]$, where $t_{k,a}$ and $t_{k,b}$ are the start time index and the end time index, respectively. For example, as shown in Fig. 2, the source signal $S_{\mathrm{D}(1)}$ (preforming operation) is dominant within Segment $\mathbf{T}_2$ with $t_{k,a} = 35$ and $t_{k,b} = 100$ obtained based on Table 1. In this situation, within the dominant time period $[t_{k,a}\ t_{k,b}]$, the ratio of two reduced sensor signals—i.e., $\tilde{X}_i(t)$ and $\tilde{X}_j(t)$, denoted as $\lambda_{i,j}(t)$—is calculated as

$$\lambda_{i,j}(t) = \frac{\tilde{X}_i(t)}{\tilde{X}_j(t)}. \tag{9}$$

Substituting Equation (8) into Equation (9):

$$\lambda_{i,j}(t) = \frac{A_{i,\mathrm{D}(k)}S_{\mathrm{D}(k)}(t) + A_{i,\mathrm{D}(1)}S_{\mathrm{D}(1)}(t) + \cdots + A_{i,\mathrm{D}(k-1)}S_{\mathrm{D}(k-1)}(t) + A_{i,\mathrm{D}(k+1)}S_{\mathrm{D}(k+1)}(t) + \cdots + A_{i,\mathrm{D}(q)}S_{\mathrm{D}(q)}(t) + \tau_i(t)}{A_{j,\mathrm{D}(k)}S_{\mathrm{D}(k)}(t) + A_{j,\mathrm{D}(1)}S_{\mathrm{D}(1)}(t) + \cdots + A_{j,\mathrm{D}(k-1)}S_{\mathrm{D}(k-1)}(t) + A_{j,\mathrm{D}(k+1)}S_{\mathrm{D}(k+1)}(t) + \ldots + A_{j,\mathrm{D}(q)}S_{\mathrm{D}(q)}(t) + \tau_j(t)} \quad (10)$$

Because the dependent source signal $S_{\mathrm{D}(k)}(t)$ is assumed to be dominant at the time interval $[t_{k,a}\ t_{k,b}]$, $S_{\mathrm{D}(k)}(t)$ provides a dominant contribution to the reduced sensor signals $\tilde{X}_h(t), h = 1, \ldots, m$; i.e., $\tilde{X}_h(t) \approx A_{h,\mathrm{D}(k)}S_{\mathrm{D}(k)}(t)$. In this case, $\lambda_{i,j}(t)$ can be further calculated as

$$\lambda_{i,j}(t) \approx \frac{A_{i,\mathrm{D}(k)}S_{\mathrm{D}(k)}(t)}{A_{j,\mathrm{D}(k)}S_{\mathrm{D}(k)}(t)} = \frac{A_{i,\mathrm{D}(k)}}{A_{j,\mathrm{D}(k)}} \quad (11)$$

Equation (11) shows that $\lambda_{i,j}(t)$ is an estimation of $A_{i,\mathrm{D}(k)}/A_{j,\mathrm{D}(k)}$. Without loss of generality, assume $A_{1,\mathrm{D}(k)} \neq 0$. Then, $\lambda_{k,1}(t)$, $k = 2, \ldots, m$, can be estimated based on Equation (11) with $j = 1$. Thus, $[1\ \lambda_{2,1}(t), \ldots, \lambda_{m,1}(t)]^{\mathrm{T}}$ is an estimation of $(1/A_{1,\mathrm{D}(k)})\mathbf{a}_{\mathrm{D}(k)}$; i.e., an estimation of the kth column of $\mathbf{A}_{\mathrm{D}}$ due to a scale factor of $A_{1,\mathrm{D}(k)}$. It needs to be noticed that the dominance of a source is a sufficient condition of Equation (11), rather than a necessary condition. However, in practice, if signal k is not dominant, there is little chance that Equation (11) is satisfied for consecutive time points. Therefore, this linear relationship property has been extensively adopted in the SCA literature to judge the dominance of a source signal, and it is also used in this article.

Generally, the source signals overlap in the time domain; thus, the sparse property of each dependent source signal does not hold in the time domain. For example, as shown in Fig. 2, the active working range of Station 2 (blocker operation) is always overlapping with that of other stations. In this case, a linear transformation is needed to satisfy the sparse property in the transformed feature domain; e.g., a frequency domain, a time/frequency domain, or a wavelet domain. Suppose $\mathbf{\Phi}$, $\mathbf{\Phi} \in \Re^{N\times L}$ is a selected linear transform matrix that transforms source signals from the time domain to a new feature domain. Based on Equation (8):

$$\tilde{X}_i(t)\mathbf{\Phi} = \sum_{k=1}^{q} A_{i,\mathrm{D}(k)}S_{\mathrm{D}(k)}(t)\mathbf{\Phi} + \tau_i(t)\mathbf{\Phi}. \quad (12)$$

Let $\tilde{F}_i(\theta) = \tilde{X}_i(t)\mathbf{\Phi}(\tilde{F}_i(\theta) \in \Re^{1\times L}$, $\theta = 1, 2, \ldots, L$, $i = 1, 2, \ldots, m$,) denote the transformed feature vector of the reduced sensor signals, and $F_k(\theta) = S_{\mathrm{D}(k)}(t)\mathbf{\Phi}$, ($\theta = 1, 2, \ldots, L$, $k = 1, 2, \ldots, q$) represents the transformed feature vector of the dependent source signals. If there exists a feature subset in the linear transform domain that the kth dependent source signal—i.e., $F_k(\theta)$—is dominant, the kth column of $\mathbf{A}_{\mathrm{D}}$ can be similarly estimated by using $F_k(\theta)$ instead of $S_{\mathrm{D}(k)}(t)$ based on Equation (11). Specifically, the time domain can be treated as a linear transform domain with Φ being chosen as an $N \times N$ identity matrix (L is equal to N). Thus, we will make no difference between the time domain and the linear transform domain in the following discussion of the article.

4.2. *A novel SCA method to estimate A_{D} using multiple linear transforms*

There are many different methods that can be used to estimate the mixing matrix $\mathbf{A}_{\mathrm{D}}$ reported in the SCA literature (Jourjine *et al*., 2000; Lewicki and Sejnowski, 2000; Bofill and Zibulevsky, 2001; Zibulevsky and Pearlmutter, 2001; Yilmaz and Rickard, 2004; Abrard and Deville, 2005; Li *et al*., 2006). However, most methods assume that in a single linear transform domain (e.g., frequency domain, time/frequency domain, or wavelet domain), all of the individual source signals have some dominant components, and this single linear transform is pre-given. In other word. each of the column vectors of matrix $\mathbf{A}_{\mathrm{D}}$ can be estimated in a single given linear transform domain. For example, Abrard and Deville (2005) assumed that all of the source signals have some dominant components in the time/frequency domain; Li *et al*. (2006) assumed that for each of the source signals there exist feature subsets in the wavelet domain where this source signal is dominant. However, in practice, it is usually unknown whether some dominant components exist in the given linear transform domain for the source signals. In addition, due to the complexity of the source signals, it is difficult to find one linear transform for which all the source signals are dominant in the specific transformed feature subsets. In this research, a novel SCA method is proposed to estimate the mixing matrix $\mathbf{A}_{\mathrm{D}}$ based on multiple linear transform domains, which does not require that all of the source signals are dominant in a given linear transform domain. In this proposed SCA method, multiple linear transform domains are checked, and a portion of columns of $\mathbf{A}_{\mathrm{D}}$ are estimated in each of these linear transform domains.

The following subsections are used to discuss the proposed SCA method; i.e., how to estimate the column vectors of matrix $\mathbf{A}_{\mathrm{D}}$ by using multiple linear transforms. Specifically, Section 4.2.1 discusses how to estimate a column vector of matrix $\mathbf{A}_{\mathrm{D}}$ based on the sampling data in a feature subset where only one source signal is dominant. In Section 4.2.2, a statistical hypothesis test is established that can be used to check whether two estimated column vectors are statistically equivalent. Finally, Section 4.2.3 provides the detailed implementation procedure for estimating matrix $\mathbf{A}_{\mathrm{D}}$ by using multiple linear transforms.

4.2.1. *Estimation of a column vector using multiple linear transforms*

Most of the SCA methods in the literature assume that all of the individual source signals have some dominant components in a single linear transform domain. These methods cannot be directly adopted for use in the case of multiple linear transforms. For this purpose, a method is developed to estimate a column vector of matrix $\mathbf{A}_\mathrm{D}$ in this subsection.

In this article, a multiple linear regression model is used to represent the relationship between $\tilde{F}_i(\theta)$, $i = 2, 3, \ldots, m$, and $\tilde{F}_1(\theta)$ in the feature subset $[\theta_{k,a}, \theta_{k,b}]$ as follows:

$$\tilde{F}_{i+1}(\theta) = r_{i+1,1}\tilde{F}_1(\theta) + \upsilon_i(\theta), \quad i = 1, 2, \ldots, m-1, \tag{13}$$

where $r_{i+1,1}$ represents the unknown parameter and $r_{i+1,1} = A_{i+1,\mathrm{D}(k)}/A_{1,\mathrm{D}(k)}$, $\upsilon_i(\theta) \in \Re^{1\times c}$ denotes the estimation error vector that is assumed to be cross-correlated but time independent and c denotes the number of potential dominant features in the feature subset $[\theta_{k,a}\ \theta_{k,b}]$; thus $c = b - a + 1$. In addition, $\upsilon_i(\theta)$ is assumed to be normally distributed with the zero mean and covariance matrix $\Sigma \in \Re^{(m-1)\times(m-1)}$. Let $\mathbf{R} = [r_{2,1}\ r_{3,1}\ \ldots\ r_{m,1}]^\mathrm{T} \in \Re^{(m-1)\times 1}$. Vector $[1\ \mathbf{R}^\mathrm{T}]^\mathrm{T}$ estimates the kth column of matrix $\mathbf{A}_\mathrm{D}$; i.e., $\mathbf{a}_{\mathrm{D}(k)}$ subjects to a scale factor $A_{1,\mathrm{D}(k)}$. Suppose in a feature subset $\theta \in [\theta_{k,a}, \theta_{k,b}]$ of a linear transformed domain the kth dependent source signal $F_k(\theta)$ is dominant. Based on Equation (11), signals $\tilde{F}_i(\theta)$ and $\tilde{F}_j(\theta)$, $i \neq j$, have a linear relationship that can be represented by a line going through the origin of the coordinates. Specifically, the scatter plot of $\tilde{F}_i(\theta)$, $i = 2, \ldots, m$, versus $\tilde{F}_1(\theta)$ in feature subset $[\theta_{k,a}, \theta_{k,b}]$ consists of $m-1$ lines going through the origin of the coordinates. Figure 4 shows a scatter plot of $\tilde{F}_i(\theta)$, $i = 2, \ldots, 5$, versus $\tilde{F}_1(\theta)$ in Segment $\mathbf{T}_{10}$; i.e., [672, 720] for the forging process shown in Fig. 1, in which the dependent source signal of the finisher operation is dominant.

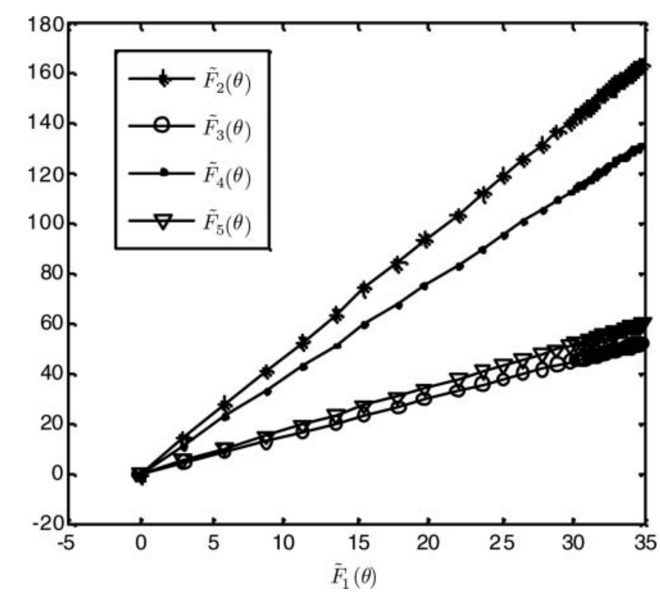

Fig. 4. Scatter plots of $\tilde{F}_i(\theta)$, $i = 2, \ldots, 5$, versus $\tilde{F}_1(\theta)$ based on the original data in segment $\mathbf{T}_{10}$ having one dominant source signal.

Let $\mathbf{Z}_i = [\,\tilde{F}_i(\theta_{k,a})\ \tilde{F}_i(\theta_{k,a+1}), \ldots, \tilde{F}_i(\theta_{k,b})]^\mathrm{T} \in \Re^{c\times 1}$ represent the transformed feature vector in the feature subset $[\theta_{k,a}, \theta_{k,b}]$ and $\mathbf{Z} = [\,\mathbf{Z}_2\ \mathbf{Z}_3, \ldots, \mathbf{Z}_m] \in \Re^{c\times(m-1)}$. In this way, the linear regression model (13) can be generally represented as a multiple linear regression model of $\mathbf{Z} = \mathbf{Z}_1\mathbf{R}^\mathrm{T} + \upsilon$, and $\upsilon = [(\upsilon_1(\theta))^\mathrm{T}\ (\upsilon_2(\theta))^\mathrm{T}, \ldots, (\upsilon_{m-1}(\theta))^\mathrm{T}] \in \Re^{c\times(m-1)}$. The maximum likelihood estimation of $\mathbf{R}$, denoted as $\hat{\mathbf{R}} = [\hat{r}_{2,1}\ \hat{r}_{3,1}\ \ldots\ \hat{r}_{m,1}]^\mathrm{T} \in \Re^{(m-1)\times 1}$, is calculated as (Lewicki and Sejnowski, 2000):

$$\hat{\mathbf{R}} = \left(\left(\mathbf{Z}_1^\mathrm{T}\mathbf{Z}_1\right)^{-1}\mathbf{Z}_1^\mathrm{T}\mathbf{Z}\right)^\mathrm{T}, \tag{14}$$

and the estimated residuals, denoted as $\hat{\upsilon} \in \Re^{c\times(m-1)}$, are obtained as

$$\hat{\upsilon} = \left[\mathbf{I} - \mathbf{Z}_1\left(\mathbf{Z}_1^\mathrm{T}\mathbf{Z}_1\right)^{-1}\mathbf{Z}_1^\mathrm{T}\right]\mathbf{Z}. \tag{15}$$

The estimate $\hat{\mathbf{R}}$ has a normal distribution with mean $\mathbf{R}$ and a covariance matrix $(\mathbf{Z}_1^\mathrm{T}\mathbf{Z}_1)^{-1}\Sigma$; and the maximum likelihood estimation of covariance matrix Σ is $\hat{\upsilon}^\mathrm{T}\hat{\upsilon}/c$.

In the SCA literature, to identify the feature subsets where only one source signal is dominant, the feature set in the linear transform domain is divided into feature subsets that do not overlap with other feature sets. In general, the division criterion depends on the selected linear transform. For example, the divided feature subsets via an STFT generally have the same cardinality in the time and frequency regions. In this article, the linear relationship between $\tilde{F}_i(\theta)$, $i = 2, 3, \ldots, m$, and $\tilde{F}_1(\theta)$ is checked in each feature subset. The problem of checking whether the response variables and the independent variable have a linear relationship has been studied in the literature. For a detailed description of this study, please refer to Cohen (2003). If the linear relationship between $\tilde{F}_i(\theta)$, $i = 2, 3, \ldots, m$, and $\tilde{F}_1(\theta)$ (modeled in Equation (13)) holds in the feature subset this indicates that one single source signal is dominant in this feature subset.

4.2.1.1. *Statistical equality testing of two estimated columns.* In practice, a source signal can have dominant components in multiple feature subsets based on which of the estimated columns correspond to the same column vector of $\mathbf{A}_\mathrm{D}$. However, due to random noise, small differences exist in these estimated columns. In this article, a statistical hypothesis testing procedure is established to test the statistical equality of two estimated columns. Since the column of $\mathbf{A}_\mathrm{D}$ is determined by vector $\mathbf{R}$, the problem of testing whether two estimated columns are statistically equivalent can be formulated as testing the statistical equality of two sample vectors—i.e., $\hat{\mathbf{R}}^k$ and $\hat{\mathbf{R}}^j$—which are estimated based on two different feature subsets. Here a superscript i is used to indicate that the ith feature subset of sampling data is used to estimate $\mathbf{R}^i$, Σ^i, and $\mathbf{Z}^i$, $i = k$ or j, $k \neq j$, which will be used in the following discussion. Since the mean vector of $\hat{\mathbf{R}}^i$ equals $\mathbf{R}^i$, the statistical testing procedure is used to test the equality of the two mean vectors; i.e., $\mathbf{R}^k$ and $\mathbf{R}^j$.

Define the test hypothesis as

$$\begin{aligned} H_0 &: \mathbf{R}^k = \mathbf{R}^j, \\ H_1 &: \mathbf{R}^k \neq \mathbf{R}^j \end{aligned} \tag{16}$$

and define the statistic Λ as

$$\Lambda = (\hat{\mathbf{R}}^k - \hat{\mathbf{R}}^j)^{\mathrm{T}} \Delta^{-1} (\hat{\mathbf{R}}^k - \hat{\mathbf{R}}^j) \tag{17}$$

where $\Delta = \Delta^k + \Delta^j \in \Re^{(m-1)\times(m-1)}$, and Δ^i is calculated as $\Delta^i = ((\hat{\boldsymbol{\upsilon}}^i)^T \hat{\boldsymbol{\upsilon}}^i / ((c^i - 1) \times (\mathbf{Z}^i)^T \mathbf{Z}^i)) \in \Re^{(m-1)\times(m-1)}$.

The decision rule to test hypothesis (16) is given in Proposition 2 and the detailed proof of Proposition 2 is given in the Appendix. The principle is to test the equality of two mean vectors without assuming the equality of the covariance matrix.

Proposition 2. *The hypothesis testing (16) is accepted if and only if*

$$\Lambda \leq \frac{\gamma(m-1)}{(\gamma - m + 2)} F_{m-1,\gamma-m+2}(\alpha), \tag{18}$$

where the degree of freedom γ is calculated as

$$\gamma = \frac{(m^2 - m)}{1/(c^1 - 1) \times \{\mathrm{tr}[(\Delta^1\Delta^{-1})^2] + [\mathrm{tr}(\Delta^1\Delta^{-1})]^2\} + 1/(c^2 - 1) \times \{\mathrm{tr}[(\Delta^2\Delta^{-1})^2] + [\mathrm{tr}(\Delta^2\Delta^{-1})]^2\}} \tag{19}$$

and $F_{i,j}(\alpha)$ is the (1 − α) percentile for F distribution with degrees i and j.

4.2.2. *Implementation procedures for estimating matrix $\mathbf{A}_D$ using multiple linear transforms*

Based on Sections 4.2.1 and 4.2.2, a systematic implementation procedure is provided in this section for estimating $\mathbf{A}_\mathrm{D}$, which is described in detail in the following.

Step 1. Initially set the estimated set of columns as $\Psi = \emptyset$.

Step 2. Transform the reduced sensor signals $\tilde{X}_i(t)$, $i = 1, \ldots, m$, into a linear transform domain based on Equation (12).

Step 3. Check the linear relationship between $\tilde{F}_i(\theta)$, $i = 2, 3, \ldots, m$, and $\tilde{F}_1(\theta)$ for each feature subset and identify the feature subsets in which only one source signal is dominant; i.e., check whether the linear relationship between $\tilde{F}_i(\theta)$, $i = 2, 3, \ldots, m$, and $\tilde{F}_1(\theta)$ in Equation (13) holds.

Step 4. For each feature subset k in which only one source signal is dominant, denoted as $[\theta_{k,a}, \theta_{k,b}]$, in the linear transform domain.

- *4.1.* Estimate the column of matrix $\mathbf{A}_\mathrm{D}$, denoted as $\hat{\mathbf{R}}^k$, based on Equation (14).
- *4.2.* If $\Psi = \emptyset$, set $\Psi = \{\hat{\mathbf{R}}^k\}$ and go to Step 5; else, do Steps 4.2.1 to 4.2.3
  - *4.2.1.* For each of the existing columns in Ψ, denoted as $\hat{\mathbf{R}}^j$, test the statistical equality of $\hat{\mathbf{R}}^k$ and $\hat{\mathbf{R}}^j$; i.e., calculate statistics Λ based on Equation (17) and test hypothesis (16) based on Proposition 2.
  - *4.2.2.* If the null hypothesis in Equation (16) is not rejected for column $\hat{\mathbf{R}}^j$ in Ψ, this indicates that $\hat{\mathbf{R}}^k$ and $\hat{\mathbf{R}}^j$ estimate the same column of $\mathbf{A}_\mathrm{D}$. In this case, an estimation with the small variance will be chosen; i.e., if condition $|\hat{\boldsymbol{\upsilon}}^k| < |\hat{\boldsymbol{\upsilon}}^j|$ holds (where $|\bullet|$ denotes the matrix determinant), substitute $\hat{\mathbf{R}}^k$ for $\hat{\mathbf{R}}^j$ in Ψ.
  - *4.2.3.* If the null hypothesis in Equation (16) is rejected for all of the elements in Ψ, this indicates that $\hat{\mathbf{R}}^k$ estimates a new column of $\mathbf{A}_\mathrm{D}$, and set $\Psi = \Psi \cup \hat{\mathbf{R}}^k$.

Step 5. If the cardinality of Ψ equals q, all of the q columns of $\mathbf{A}_\mathrm{D}$ have been found; otherwise, continuously choose a new linear transform Φ and go back to Step 2.

The procedure will stop when achieving either of two situations: (i) all of the q columns of matrix $\mathbf{A}_\mathrm{D}$ have been successfully estimated; or (ii) it is impossible to obtain matrix $\mathbf{A}_\mathrm{D}$ using all the employed linear transforms.

5. Case study

The progressive forging process, as shown in Fig. 1, is used as a case study to demonstrate the developed method. In this case study, a minimum number of five mixed sensor signals have to be collected to allow estimation of the individual forces of the dies. In order to demonstrate the generality of the proposed method and to show that it is applicable to an arbitrary sensing system, the following arbitrary mixing matrices $\mathbf{A}_\mathrm{I}$ and $\mathbf{A}_\mathrm{D}$ are simulated, where each column $\mathbf{A}_\mathrm{I}$ and $\mathbf{A}_\mathrm{D}$ is normalized such that $(\mathbf{a}_{\mathrm{I}(j)})^\mathrm{T}\mathbf{a}_{\mathrm{I}(j)} = 1$ and $(\mathbf{a}_{\mathrm{D}(i)})^\mathrm{T}\mathbf{a}_{\mathrm{D}(i)} = 1$.

$$\mathbf{A}_\mathrm{I} = \begin{bmatrix} 0.55 & 0.57 \\ 0.21 & 0.67 \\ 0.48 & 0.26 \\ 0.53 & 0.44 \\ 0.38 & 0.65 \end{bmatrix}, \quad \mathbf{A}_\mathrm{D} = \begin{bmatrix} 0.29 & 0.59 & 0.15 \\ 0.13 & 0.29 & 0.72 \\ 0.71 & 0.54 & 0.23 \\ 0.35 & 0.25 & 0.58 \\ 0.46 & 0.46 & 0.26 \end{bmatrix}.$$

The sensor noise $\boldsymbol{\varepsilon}$ is assumed to follow a multivariate normal distribution $MN(0, \zeta^2\mathbf{I})$, where $\mathbf{I}$ is an $m \times m$ identity matrix, and ζ^2 is set to 0.1. Figure 5 illustrates the simulated mixed sensor signals $\mathbf{X}(t)$.

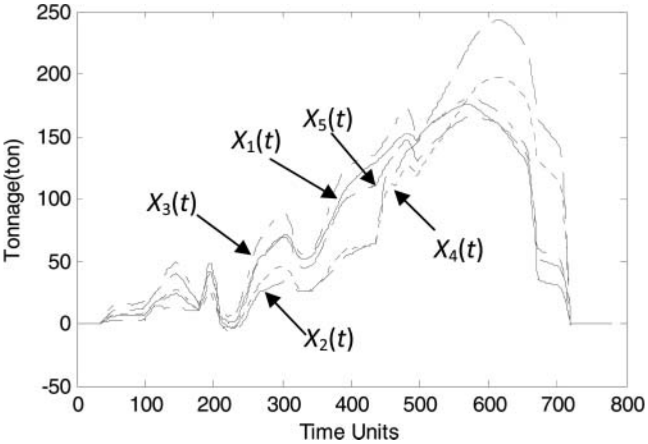

Fig. 5. Simulated mixed sensor signals $\mathbf{X}(t)$.

Table 2. Energy ratio $\eta_{i,j}$ calculated in active working range $\mathbf{L}_{\mathrm{I}(j)} \equiv [l_{\mathrm{I}(j),0}, l_{\mathrm{I}(j),1}], j = 1, 2$

	i				
$\eta_{i,j}$	*1*	*2*	*3*	*4*	*5*
$\mathbf{L}_{I(1)} = [196, 353]$	0.97	0.60	0.24	0.08	0.10
$\mathbf{L}_{I(2)} = [100, 268]$	N/A	0.96	0.05	0.25	0.42

5.1. *Estimation of independent source signals*

The ICA method is first applied to the mixed sensor signals $\mathbf{X}(t)$ to estimate the independent source signals $\mathbf{S}_{\mathrm{I}}(t)$. Figure 6 shows the estimated signals after applying the ICA method; i.e., $Y_i(t), i = 1, 2, \ldots, 5$. To identify the estimated independent source signals, the algorithm in Section 3.2 was applied and Table 2 shows energy ratio $\eta_{i,j}$ calculated in time period $\mathbf{L}_{\mathrm{I}(j)} \equiv [l_{\mathrm{I}(j),0}, l_{\mathrm{I}(j),1}], j = 1, 2$, based on Equation (6). In the first step of the algorithm, $Y_1(t)$ was chosen to estimate $S_{\mathrm{I}(1)}(t)$ since $\eta_{1,1}$ has the maximum value for the active working range $\mathbf{L}_{\mathrm{I}(1)} = [196, 353]$. Similarly, $Y_2(t)$ was chosen to estimate $S_{\mathrm{I}(2)}(t)$ in the second step. Note that because $Y_1(t)$ was already chosen at the first step, $\eta_{2,1}$ will not be calculated for $\mathbf{L}_{\mathrm{I}(2)}$.

Thus, $Y_1(t)$ and $Y_2(t)$ estimate the independent source signals $S_{\mathrm{I}(1)}(t)$ and $S_{\mathrm{I}(2)}(t)$, generated from station 4 and station 5, respectively, while $Y_3(t)$, $Y_4(t)$, and $Y_5(t)$ estimate the linear combination of dependent source signals. This can also be seen in Fig. 6.

Next, the impacts of independent source signals—i.e., $S_{\mathrm{I}(1)}(t)$ and $S_{\mathrm{I}(2)}(t)$—on the mixed sensor signals were eliminated, and the reduced sensor signals, $\tilde{X}_i(t), i = 1, 2, \ldots, 5$, were calculated based on Corollary 1.

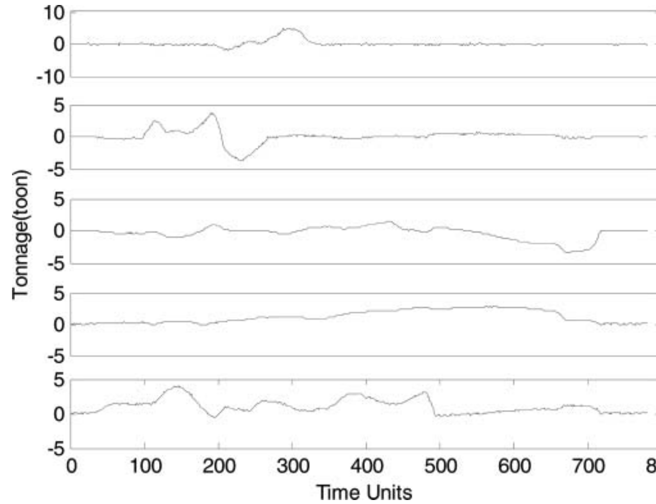

Fig. 6. The estimated signals after applying the ICA method; i.e., $Y_i\,(t), i = 1, 2, \ldots, 5$.

5.2. *Estimation of dependent source signals*

Based on the implementation procedures given in Section 4.2.3, the SCA method was used to obtain the estimated dependent source signals $\hat{\mathbf{S}}_{\mathrm{D}}(t)$. In the first step, the time domain was directly considered with $\mathbf{\Phi}$ being chosen as an $N \times N$ identity matrix. Since we only needed to estimate the dependent source signals, the data segmentation needed to be re-conducted based on only those dependent source signals; i.e., only based on the active working ranges of stations 1 to 3 to divide the whole cycle signals. The resultant seven new data segments were denoted as $\mathbf{T}_{\mathrm{D}(k)}$, $(k = 1, 2, \ldots, 7)$, and their start and end time indexes are given in Table 3. The relationship between these seven new data segments and the original 11 data segments ($\mathbf{T}_{\mathrm{k}}$, $k = 1, 2, \ldots, 11$), as shown in Fig. 2, are also shown in Table 3.

Since the sensor signals in segments $\mathbf{T}_{\mathrm{D}(1)}$ and $\mathbf{T}_{\mathrm{D}(7)}$ only consist of noises, $\mathbf{T}_{\mathrm{D}(1)}$ and $\mathbf{T}_{\mathrm{D}(7)}$ will not be further considered in estimating the dependent source signals. Figure 7 shows the scatter plots of $\tilde{F}_i(\theta)$, $(i = 2, 3, 4, 5)$ versus $\tilde{F}_1(\theta)$ in each data segment $\mathbf{T}_{\mathrm{D}(k)}$ $(k = 2, 3, \ldots, 6)$. From Fig. 7, it can be seen that the linear relationship between $\tilde{F}_i(\theta)$ and $\tilde{F}_1(\theta)$ in Equation (13) is held in two segments $\mathbf{T}_{\mathrm{D}(2)}$ and $\mathbf{T}_{\mathrm{D}(6)}$. The estimation of $\mathbf{R}$ in data segments $\mathbf{T}_{\mathrm{D}(2)}$ and $\mathbf{T}_{\mathrm{D}(6)}$, denoted as $\hat{\mathbf{R}}^1$ and $\hat{\mathbf{R}}^2$, respectively, were then calculated based on Equation (14) using the original sensing signal in segments $\mathbf{T}_{\mathrm{D}(2)}$ and $\mathbf{T}_{\mathrm{D}(6)}$, respectively. The corresponding estimation results are $\hat{\mathbf{R}}^1 = [0.50, 0.92, 0.43, 0.78]^{\mathrm{T}}$ and $\hat{\mathbf{R}}^2 = [4.71, 1.50, 3.78, 1.72]^{\mathrm{T}}$. Next, to test whether

Table 3. Time periods $\mathbf{T}_{\mathrm{D}(k)}, k = 1, 2, \ldots, 7$

$\boldsymbol{T}_{\mathrm{D}(1)} = \boldsymbol{T}_1$	$\boldsymbol{T}_{\mathrm{D}(2)} = \boldsymbol{T}_2 + \boldsymbol{T}_3$	$\boldsymbol{T}_{\mathrm{D}(3)} = \boldsymbol{T}_4 + \boldsymbol{T}_5 + \boldsymbol{T}_6 + \boldsymbol{T}_7$	$\boldsymbol{T}_{\mathrm{D}(4)} = \boldsymbol{T}_8$	$\boldsymbol{T}_{\mathrm{D}(5)} = \boldsymbol{T}_9$	$\boldsymbol{T}_{\mathrm{D}(6)} = \boldsymbol{T}_{10}$	$\boldsymbol{T}_{\mathrm{D}(7)} = \boldsymbol{T}_{11}$
[0, 35]	[35, 179]	[179, 436]	[436, 494]	[494, 672]	[672, 720]	[720, 780]

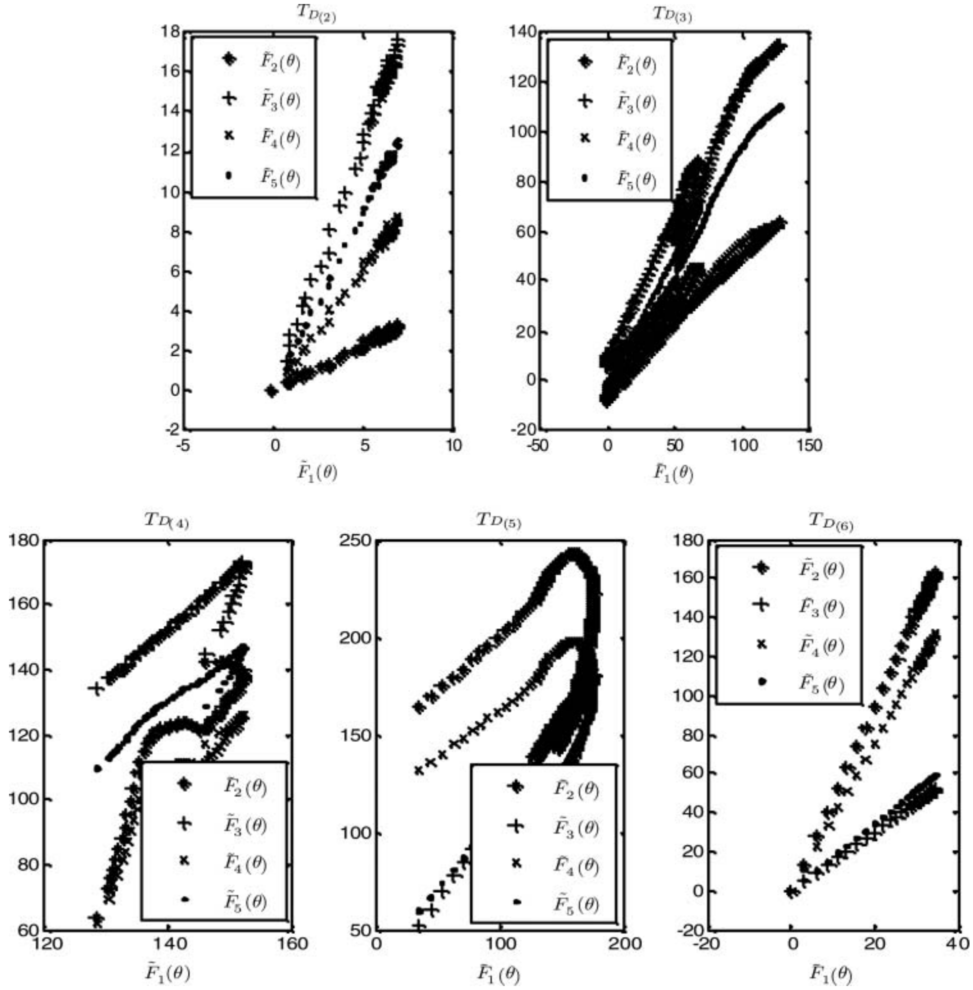

Fig. 7. Scatter plots of $\tilde{F}_i(\theta)$, $i = 2, 3, 4, 5$, versus $\tilde{F}_1(\theta)$ in time periods $\mathbf{T}_{D(i)}$, $i = 2, 3, \ldots, 6$.

$[1\ (\hat{\mathbf{R}}^1)^T]^T$ and $[1\ (\hat{\mathbf{R}}^2)^T]^T$ estimate the same column of matrix $\mathbf{A}_D$, statistic Λ and the degree of freedom γ were calculated based on Equation (17) and Equation (19), respectively. Since

$$\Lambda = 35987 > \frac{15.86 \times (5-1)}{(15.86 - 5 + 2)} F_{5-1,15.86-5+2}(0.05) = 15.74,$$

based on Proposition 2, we rejected the null hypothesis in Equation (16); i.e., $[1\ (\hat{\mathbf{R}}^1)^T]^T$ and $[1\ (\hat{\mathbf{R}}^2)^T]^T$ estimate different columns of $\mathbf{A}_D$. Thus, we get $\Psi = \{\hat{\mathbf{R}}^1, \hat{\mathbf{R}}^2\}$.

Since the cardinality of Ψ is equal to two, which is less than the number of dependent source signals, it is clear that not all of the three columns of matrix $\mathbf{A}_D$ can be estimated in the time domain. In other words, only two dependent source signals have dominant components in the time domain. This can be verified by reference to Fig. 2, in which the source signal generated by the preforming operation and that generated by the finisher operation are dominant in data segment $\mathbf{T}_{D(2)}$ (i.e., $\mathbf{T}_2 + \mathbf{T}_3$) and $\mathbf{T}_{D(6)}$ (i.e., $\mathbf{T}_{10}$), respectively, after removing independent source signals.

To estimate the remaining column of matrix $\mathbf{A}_D$, we used a linear transform based on Equation (12) to transform the reduced sensor signals $\tilde{X}_i(t)$, $i = 1, 2, \ldots, 5$, into a feature domain. In this case study, Φ was selected as the STFT with the Hamming window (Gaeta and Lacoume, 1990). Each data segment $\mathbf{T}_{D(i)}$, $(i = 2, 3, \ldots, 6)$ was further divided into seven sub-periods using the default parameters of the MATLAB program (Krauss *et al.*, 1993). Since the features of the reduced sensor signals in the STFT domain—i.e., $\tilde{F}_i(\theta)$, $i = 1, 2, \ldots, 5$—consist of complex numbers, the modulus of $\tilde{F}_i(\theta)$, denoted as $|\tilde{F}_i(\theta)|$, was used. Figure 8 illustrates the scatter plot of $|\tilde{F}_i(\theta)|$, $i = 2, 3, 4, 5$, versus $|\tilde{F}_1(\theta)|$ in the first and seventh sub-periods of the data segment $\mathbf{T}_{D(3)}$. It can be seen that the linear relationship between $|\tilde{F}_i(\theta)|$ and $|\tilde{F}_1(\theta)|$ in Equation (13) holds in the seventh sub-period of segment $\mathbf{T}_{D(3)}$. Using the sample points in the seventh sub-period of $\mathbf{T}_{D(3)}$ in the STFT domain, $\mathbf{R}^3$ is estimated as $\hat{\mathbf{R}}^3 = [0.47,\ 1.92,\ 0.93,\ 1.42]^T$.

Then, the hypothesis test (16) was used to test whether $\hat{\mathbf{R}}^3$ was statistically equivalent to $\hat{\mathbf{R}}^1$ or $\hat{\mathbf{R}}^2$. Since the hypothesis test (16) was rejected for both situations, it is apparent

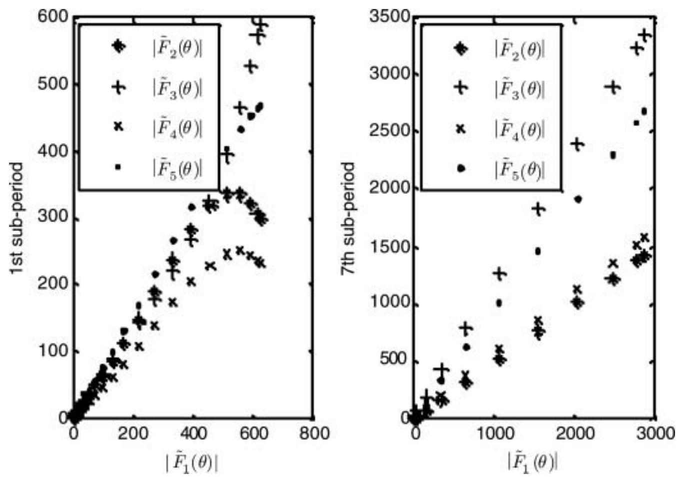

Fig. 8. Scatter plot of $|\tilde{F}_i(\theta)|$, $i = 2, 3, 4, 5$, versus $|\tilde{F}_1(\theta)|$ in the first and seventh sub-periods of $\mathbf{T}_{D(3)}$ in the STFT domain.

that $[\,1\ (\hat{\mathbf{R}}^3)^T\,]^T$ estimates a new column of matrix $\mathbf{A}_D$. Let $\hat{\mathbf{a}}_{D(i)} = [\,1\ (\hat{\mathbf{R}}^i)^T\,]^T$, $i = 1, 2, 3$, $\hat{\mathbf{a}}_{D(i)}$ estimates one of the three columns of matrix $\mathbf{A}_D$. Thus, the estimated matrix of $\mathbf{A}_D$ is given as $\hat{\mathbf{A}}_D \equiv [\,\hat{\mathbf{a}}_{D(1)},\ \hat{\mathbf{a}}_{D(2)},\ \hat{\mathbf{a}}_{D(3)}\,]$. After normalizing $\hat{\mathbf{A}}_D$ so that $(\hat{\mathbf{a}}_{D(i)})^T\hat{\mathbf{a}}_{D(i)} = 1$, the resultant matrix, denoted as $\hat{\mathbf{A}}'_D$, could be written as

$$\hat{\mathbf{A}}'_D = \begin{bmatrix} 0.28 & 0.15 & 0.56 \\ 0.13 & 0.72 & 0.27 \\ 0.72 & 0.23 & 0.57 \\ 0.34 & 0.58 & 0.27 \\ 0.52 & 0.26 & 0.48 \end{bmatrix}.$$

After obtaining $\hat{\mathbf{A}}'_D$, the estimated source signals could be calculated by using the least squares method; i.e., $\hat{\mathbf{S}}_D(t) = ((\hat{\mathbf{A}}'_D)^T\hat{\mathbf{A}}'_D)^{-1}(\hat{\mathbf{A}}'_D)^T\tilde{\mathbf{X}}(t)$. The resultant signals obtained using the proposed SCA method are shown in Fig. 9, in which the solid curves represent the true dependent source signals—i.e., $S_{D(1)}(t)$, $S_{D(2)}(t)$, and $S_{D(3)}(t)$—and the dashed curves represent the estimated dependent source signals. From Fig. 9, it can be seen that the SCA estimation method is able to closely reproduce the data.

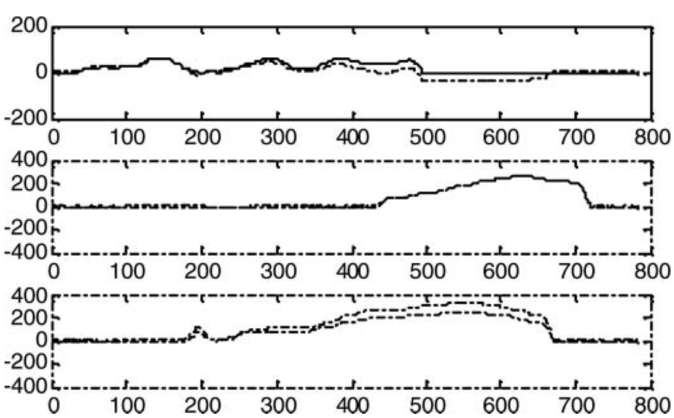

Fig. 9. Results of the SCA method.

6. Conclusions

This article proposes a method for separating general mixed source signals, which may include both independent and dependent signals. The developed estimation method consists of two major steps that combine the ICA and SCA methods. In the first step, the independent source signals are estimated based on the ICA method. Although the assumption required by the ICA method is not valid for all source signals, it has been proved in this article that the estimated signals after using the ICA method surely include the estimated independent source signals. In the second step, the impacts of the independent source signals are eliminated from the mixed sensor signals. The remaining sensing signals are then used to estimate the dependent source signals based on the SCA method.

Different from the existing literature, this article developed an SCA method to estimate the dependent source signals without requiring the assumption that every dependent source signal has dominant components by using a single given linear transform. For this purpose, a statistical testing method is proposed to check whether two estimated columns of the mixing matrix are statistically equivalent under multiple linear transforms. A case study on a forging process is conducted to demonstrate the effectiveness of the developed method.

In the future, general SPC methods will be applied to monitor the estimated individual source signals, which can provide explicit diagnostic information for individual operations to enhance the system's diagnostic abilility. Another future task will be to extend the developed separation method to a general case in which there may exist non-linear relationships between source signals and mixed sensor measurements.

Acknowledgements

This research is jointly supported by the NSF under grant DMI-0541750, Michigan 21st Century Jobs Fund F015886 and the Forging Industrial Educational & Research Foundation. The authors would also like to thank OG Technologies and FormTech for their help and support during the tests and data collection.

References

Abrard, F. and Deville, Y. (2005) A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources. *Signal Processing*, **85**, 1389–1403.

Blanchard, G., Kawanabe, M., Sugiyama, M., Spokoiny, V. and Müller, K. (2006) In search of non-Gaussian components of a high-dimensional distribution. *The Journal of Machine Learning Research*, **7**, 247–282.

Bofill, P. and Zibulevsky, M. (2001) Underdetermined blind source separation using sparse representations. *Signal Processing*, **81**, 2353–2362.

Chiu, K., Liu, Z. and Xu, L. (2003) A statistical approach to testing mutual independence of ICA recovered sources, in *Proceedings of the Fourth International Symposium on Independent Component Analysis and Blind Signal Separation*, pp. 751–756.

Cohen, J. (2003) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Lawrence Erlbaum, Mahwah, NJ.

Comon, P. (1994) Independent component analysis, a new concept. *Signal Processing*, **36**, 287–314.

Gaeta, M. and Lacoume, J.L. (1990) Source separation without *a priori* knowledge: the maximum likelihood solution. in *Proceedings of the Fifth European Signal Processing Conference*, pp. 621–624.

Herault, J. and Jutten, C. (1986) Space or time adaptive signal processing by neural network models, in *Proceedings of the AIP Conference*, vol. 151, pp. 206–211.

Hyvarinen, A. and Oja, E. (1997) A fast fixed-point algorithm for independent component analysis. *Neural Computation*, **9**, 1483–1492.

Hyvrinen, A., Karhunen, J. and Oja, E. (2001) *Independent Component Analysis*, John Wiley & Sons, New York, NY.

Jin, J. (2004) Individual station monitoring using press tonnage sensors for multiple operation stamping processes. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **126**, 83–90.

Jin, J. and Shi, J. (2005) Press tonnage signal decomposition and validation analysis for transfer or progressive die processes. *Journal of Manufacturing Science and Engineering*, **127**(1), 231–235.

Jourjine, A., Rickard, S. and Yilmaz, O. (2000) Blind separation of disjoint orthogonal signals: demixing *N* sources from 2 mixtures, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, IEEE Press, Piscataway, NJ, pp. 2985–2988.

Karvanen, J. (2005) A resampling test for the total independence of stationary time series: application to the performance evaluation of ICA algorithms. *Neural Processing Letters*, **22**, 311–324.

Krauss, T.P., Shure, L. and Little, J.N. (1993) *Signal Processing Toolbox for Use with MATLAB*, The MathWorks, Natick, MA.

Krishnamoorthy, K. and Yu, J. (2004) Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem. *Statistics and Probability Letters*, **66**, 161–169.

Lewicki, M.S. and Sejnowski, T.J. (2000) Learning overcomplete representations. *Neural Computation*, **12**, 337–365.

Li, Y., Amari, S., Cichocki, A., Ho, D.W.C. and Xie, S. (2006) Underdetermined blind source separation based on sparse representation. *IEEE Transactions on Signal Processing*, **54**, 423–437.

Skitovich, V. (1953) On a property of a normal distribution, *Doklady Academii Nauk SSSR*, **89**, 217–219.

Skitovich, V. (1954) Linear forms of independent random variables and the normal distribution law. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, **18**, 185–200.

Yilmaz, O. and Rickard, S. (2004) Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, **52**, 1830–1847.

Zibulevsky, M. and Pearlmutter, B.A. (2001) Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, **13**, 863–882.

Appendix

Proof of Proposition 1. It is assumed that the independent/dependent source signals are non-Gaussian and that the sensor noise—i.e., $\varepsilon(t)$—is a white noise and independent of the source signals; i.e., $\varepsilon(t) \sim MN(0, \zeta^2\mathbf{I})$. The method in Blanchard *et al.* (2006) can be used to obtain the subspace that only contains the non-Gaussian source signals based on the mixed sensor measurements $\mathbf{X}(t)$. Therefore, the sensor noises $\varepsilon(t)$ will be considered in the following discussion.

Substituting Equation (3) into Equation (2) leads to

$$\mathbf{Y}(t) = \mathbf{W} \times [\mathbf{A}_\mathrm{I}\, \mathbf{A}_\mathrm{D}] \times \begin{bmatrix} \mathbf{S}_\mathrm{I}(t) \\ \mathbf{S}_\mathrm{D}(t) \end{bmatrix} \tag{A1}$$

Let

$$\mathbf{Y}(t) = \begin{bmatrix} \mathbf{Y}_\mathrm{I}(t) \\ \mathbf{Y}_\mathrm{D}(t) \end{bmatrix},$$

where $\mathbf{Y}_\mathrm{I}(t) \in \Re^{p\times N}$ and $\mathbf{Y}_\mathrm{D}(t) \in \Re^{q\times N}$; and let

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix},$$

where $\mathbf{W}_1 \in \Re^{p\times m}$ and $\mathbf{W}_2 \in \Re^{q\times m}$, such that $\mathbf{Y}_\mathrm{I}(t)$ and $\mathbf{Y}_\mathrm{D}(t)$ estimate independent source signals $\mathbf{S}_\mathrm{I}(t)$ and dependent source signals $\mathbf{S}_\mathrm{D}(t)$, respectively. $\mathbf{Y}_\mathrm{I}(t)$ and $\mathbf{Y}_\mathrm{D}(t)$ can be calculated based on Equation (A1) as

$$\begin{aligned} \mathbf{Y}_\mathrm{I}(t) &= \mathbf{W}_1\mathbf{A}_\mathrm{I}\mathbf{S}_\mathrm{I}(t) + \mathbf{W}_1\mathbf{A}_\mathrm{D}\mathbf{S}_\mathrm{D}(t) \\ \mathbf{Y}_\mathrm{D}(t) &= \mathbf{W}_2\mathbf{A}_\mathrm{I}\mathbf{S}_\mathrm{I}(t) + \mathbf{W}_2\mathbf{A}_\mathrm{D}\mathbf{S}_\mathrm{D}(t) \end{aligned} \tag{A2}$$

Next we prove $\mathbf{W}_1\mathbf{A}_\mathrm{D} = 0$ and $\mathbf{W}_2\mathbf{A}_\mathrm{I} = 0$. First, we prove $\mathbf{W}_2\mathbf{A}_\mathrm{I} = 0$. The estimated result of the ICA method—i.e., $\mathbf{Y}_\mathrm{I}(t)$ and $\mathbf{Y}_\mathrm{D}(t)$—are mutually independent (Hyvrinen *et al.*, 2001). Suppose $\mathbf{W}_2\mathbf{A}_\mathrm{I}$ is not equal to zero. By arranging the order of columns appropriately while keeping track of the column indices, a reduced row echelon form of $\mathbf{W}_2\mathbf{A}_\mathrm{I}$ can be generally written in the form of

$$\left[\begin{array}{c|c} \mathbf{I}_r & \mathbf{D} \\ \mathbf{0} & \mathbf{0} \end{array}\right],$$

where $r = \mathrm{rank}(\mathbf{W}_2\mathbf{A}_\mathrm{I})$ and $\mathbf{I}_r$ is an $r \times r$ identity matrix that is a linear transformation of the submatrix of $\mathbf{W}_2\mathbf{A}_\mathrm{I}$. Based on the multivariate Skitovitch–Darmois theorem (Skitovich, 1953, 1954), the signals in $\mathbf{S}_\mathrm{I}(t)$ corresponding to $\mathbf{I}_r$ are Gaussian, which contradicts the assumption that $\mathbf{S}_\mathrm{I}(t)$ are non-Gaussian. Thus, $\mathbf{W}_2\mathbf{A}_\mathrm{I} = 0$. Similarly, it can be proved that $\mathbf{W}_1\mathbf{A}_\mathrm{D} = 0$. Thus, we get $\mathbf{Y}_\mathrm{I}(t) = \mathbf{W}_1\mathbf{A}_\mathrm{I}\mathbf{S}_\mathrm{I}(t)$ and $\mathbf{Y}_\mathrm{D}(t) = \mathbf{W}_2\mathbf{A}_\mathrm{D}\mathbf{S}_\mathrm{D}(t)$. Let $\hat{\mathbf{S}}_{\mathrm{D}'}(t) = \mathbf{Y}_\mathrm{D}(t)$ and $\mathbf{D} = \mathbf{W}_2\mathbf{A}_\mathrm{D}$, $\hat{\mathbf{S}}_{\mathrm{D}'}(t)$ estimates $\mathbf{D}\mathbf{S}_\mathrm{D}(t)$.

Since it is assumed that all of the source signals $\mathbf{S}_\mathrm{I}(t)$ are independent, based on Skitovitch–Darmois theory (Skitovich, 1953, 1954) and the ICA theory (Hyvrinen *et al.* 2001), $\mathbf{W}_1\mathbf{A}_\mathrm{I}$ is a diagonal matrix due to a permutation of the columns. As a result, $\mathbf{Y}_\mathrm{I}(t)$ estimates the source signals $\mathbf{S}_\mathrm{I}(t)$ due to scales and a permutation; i.e., $\mathbf{Y}_\mathrm{I}(t) = \hat{\mathbf{S}}_\mathrm{I}(t)$.■

Proof of Corollary 1. From Equation (2), the mixed sensor signals can be further calculated as

$$\mathbf{X}(t) = [\mathbf{A}_\mathrm{I}\mathbf{A}_\mathrm{D}] \times \begin{bmatrix} \mathbf{S}_\mathrm{I}(t) \\ \mathbf{D}^{-1}\mathbf{D}\mathbf{S}_\mathrm{D}(t) \end{bmatrix} + \varepsilon(t)$$

$$= [\mathbf{A}_\mathrm{I}\mathbf{A}_\mathrm{D}\mathbf{D}^{-1}] \times \begin{bmatrix} \mathbf{S}_I(t) \\ \mathbf{D}\mathbf{S}_\mathrm{D}(t) \end{bmatrix} + \varepsilon(t).$$

Based on Proposition 1, the estimated signals obtained using the ICA method are $\mathbf{Y}(t) = \hat{\mathbf{S}}_\mathrm{I}(t) \cup \hat{\mathbf{S}}_{\mathrm{D}'}(t)$, where $\hat{\mathbf{S}}_\mathrm{I}(t)$ estimates the independent source signals $\mathbf{S}_\mathrm{I}(t)$ and $\hat{\mathbf{S}}_{\mathrm{D}'}(t)$ estimates the linear combinations of dependent source signals; i.e., $\mathbf{D}\mathbf{S}_\mathrm{D}(t)$. Suppose $\mathbf{U} = (\mathbf{W}^\mathrm{T}\mathbf{W})^{-1}\mathbf{W}^\mathrm{T}$; let $\mathbf{U}_\mathrm{D} \in \Re^{m\times q}$ be a matrix composed of the column vectors of $\mathbf{U}$, which corresponds to the source signals in $\hat{\mathbf{S}}_{\mathrm{D}'}(t)$ and let $\mathbf{U}_\mathrm{I} \in \Re^{m\times p}$ be a matrix composed of the column vectors of $\mathbf{U}$ that correspond to the source signals in $\hat{\mathbf{S}}_\mathrm{I}(t)$. Based on the ICA method, $\mathbf{U}_\mathrm{I}$ estimates $\mathbf{A}_\mathrm{I}$. As a consequence, $\mathbf{U}_\mathrm{I}$ $\mathbf{U}_\mathrm{I}\hat{\mathbf{S}}_\mathrm{I}(t)$ estimates $\mathbf{A}_\mathrm{I}\mathbf{S}_\mathrm{I}(t)$; i.e., the impact of the independent source signals on the mixed sensor signals.

Also, we have that:

$$\mathbf{U}_\mathrm{D}\hat{\mathbf{S}}_{\mathrm{D}'}(t) + \mathbf{U}_\mathrm{I}\hat{\mathbf{S}}_\mathrm{I}(t) = \mathbf{U}\mathbf{Y}(t) = (\mathbf{W}^\mathrm{T}\mathbf{W})^{-1}\mathbf{W}^\mathrm{T}\mathbf{Y}(t). \quad \text{(A3)}$$

Substituting Equation (3) into Equation (A3) results in

$$\mathbf{U}_\mathrm{D}\hat{\mathbf{S}}_{\mathrm{D}'}(t) + \mathbf{U}_\mathrm{I}\hat{\mathbf{S}}_\mathrm{I}(t) = (\mathbf{W}^\mathrm{T}\mathbf{W})^{-1}\mathbf{W}^\mathrm{T}\mathbf{W}\mathbf{X}(t) = \mathbf{X}(t). \quad \text{(A4)}$$

Thus, $\tilde{\mathbf{X}}(t) = \mathbf{X}(t) - \mathbf{U}_\mathrm{I}\hat{\mathbf{S}}_\mathrm{I}(t) = \mathbf{U}_\mathrm{D}\hat{\mathbf{S}}_{\mathrm{D}'}(t)$ ■

Proof of Proposition 2. Because $\hat{\mathbf{R}}^i$ can be estimated based on feature sets in different linear transform domains, the covariance matrix of $\upsilon^i(\theta)$—i.e., Σ^i—in Equation (13) may be different. Thus, hypothesis testing Equation (16) is to test the equality of the two mean vectors—i.e., $\mathbf{R}^k$ and $\mathbf{R}^j$—under heteroscedastic dispersion matrices.

Let $\tilde{\Delta}^i = (\hat{\mathbf{\upsilon}}^i)^\mathrm{T}\hat{\mathbf{\upsilon}}^i \in \Re^{(m-1)\times(m-1)}$. $\tilde{\Delta}^i$ and $\hat{\mathbf{R}}^i$ have the following properties (Krishnamoorthy and Yu, 2004): $\tilde{\Delta}^i$ has a distribution of $W_{m-1}(c^i - 1, \Sigma^i)$, where $W_k(a, \mathbf{B})$ denotes the k-dimensional Wishart distribution with the degree of a and scale matrix $\mathbf{B}$. $\hat{\mathbf{R}}^i$ has a normal distribution with mean $\mathbf{R}$ and covariance matrix $((\mathbf{Z}_1^i)^\mathrm{T}\mathbf{Z}_1^i)^{-1}\Sigma^i$. In addition, $\hat{\mathbf{R}}^j$, $\hat{\mathbf{R}}^k$, $\tilde{\Delta}^j$, and $\tilde{\Delta}^k$, $j \neq k$, are independent variables.

Let $\Delta^i = (\tilde{\Delta}^i/(c^i - 1)(\mathbf{Z}^i)^T\mathbf{Z}^i) \in \Re^{(m-1)\times(m-1)}$. Δ^ihas a Wishart distribution; i.e., $W_{n-1}(c^i - 1, (\Sigma^i/(c^i - 1)$ $(\mathbf{Z}^i)^T\mathbf{Z}^i))$. Let $\Delta = \Delta^k + \Delta^j \in \Re^{(n-1)\times(n-1)}$ and

$$\tilde{\Sigma} = \frac{\Sigma^j}{(\mathbf{Z}^j)^\mathrm{T}\mathbf{Z}^j} + \frac{\Sigma^k}{(\mathbf{Z}^k)^\mathrm{T}\mathbf{Z}^k}.$$

Based on Herault and Jutten (Krishnamoorthy and Yu, 2004), $\tilde{\Sigma}^{-1/2}\Delta\tilde{\Sigma}^{-1/2}$ has a distribution of $W_{m-1}(\gamma, \mathbf{I}_{m-1}/\gamma)$ approximately, where $\mathbf{I}_{m-1} \in \Re^{(m-1)\times(m-1)}$ is an identity matrix and γ is calculated in Equation (19). As a result, $(\hat{\mathbf{R}}^k - \hat{\mathbf{R}}^j)^\mathrm{T}\Delta^{-1}(\hat{\mathbf{R}}^k - \hat{\mathbf{R}}^j)$ has an approximate distribution of $(\gamma(m-1)/(\gamma - m + 2))F_{m-1,\gamma-m+2}(\alpha)$. Thus, hypothesis testing Equation (16) is accepted if and only if Equation (18) holds. ■

Biographies

Qingyu Yang received B.S. and M.S. degrees in Automatic Control and Intelligent Systems from the Univeristy of Science and Technology of China in 2000 and 2003, respectively; an M.S. degree in Statistics; and a Ph.D. degree in Industrial Engineering, both from the University of Iowa in 2007 and 2008, respectively. Currently, he is an Assistant Professor in the Department of Industrial and System Engineering at Wayne State University. His research interests include statistical data analysis, complex system modeling, and system informatics. He is a member of INFORMS and IIE.

Jionghua (Judy) Jin is a Professor in the Department of Industrial and Operations Engineering at the University of Michigan. She received her Ph.D. degree from the University of Michigan in 1999. Her recent research focuses on data fusion for improving complex system operations, which includes variation reduction via condition monitoring, diagnosis and control, knowledge discovery, and decision making. Her research integrates applied statistics, signal processing, reliability engineering, system control, and decision-making theory. She has received numerious awards including the NSF CAREER Award in 2002 and the PECASE Award in 2004 and six Best Paper Awards between 2005 and 2010. She is a member of ASME, ASQ, IEEE, IIE, INFORMS, and SME.