IIE Transactions (2007) **39**, 251–263

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170600735546

Pattern matching for variation-source identification in manufacturing processes in the presence of unstructured noise

ZHIGUO LI[1], SHIYU ZHOU[1,*] and YU DING[2]

[1] *Department of Industrial and Systems Engineering, University of Wisconsin – Madison, Madison, WI 53706, USA*
E-mail: szhou@engr.wisc.edu
[2] *Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843, USA*

Received July 2005 and accepted February 2006

Variation-source identification has received considerable attention from the manufacturing quality improvement community. One widely used method is based on a pattern matching procedure, which identifies process faults by comparing the fault symptom, which is the principal eigenvector of the covariance matrix of the quality measurement, with fault signatures. The presence of unstructured noise as well as the uncertainty due to sampling will cause the direction of the fault symptom to deviate from the corresponding fault signature. The influences of these two effects on pattern matching procedures have previously been studied separately, by assuming either the absence of unstructured noise or the availability of large samples. This paper developes a robust pattern matching procedure that considers both effects simultaneously. Using a machining process as an illustrative example, the paper demonstrates that previous pattern matching procedures can have a remarkably low identification capability when the assumptions are not strictly satisfied. By contrast, our proposed method is more robust, maintaining a good identification probability, and would be a preferable tool for root-cause identification in manufacturing quality improvement.

Keywords: Statistical pattern matching, statistical testing, unstructured noise, variation source identification

1. Introduction

Process monitoring and control technology, which focuses on the detection, identification, diagnosis, and elimination of process faults, can help a company maintain on edge in today's highly competative market place. This is because it can help to reduce process downtime, and hence, the operation costs. The rapid advances in sensing and information technology that are currently being made mean that a large amount of data is readily available that requires process control methodologies to be developed for its interpretation.

Statistical Process Control (SPC) (Montgomery (2005) and the references therein) is the primary tool used in practice to improve the quality of manufacturing processes. SPC methods compare the statistical distribution of a process output at normal (in control) working conditions with that at current working conditions. If a large disparity is found, then an alarm is signaled to indicate an abnormal (out of control) condition. However, SPC is purely a statistical technique that is able to detect a departure from normal conditions but is unable to pin down the process fault that caused the alarm. This process fault is often called the root cause of the alarm by practitioners. The job of root-cause identification is actually left to plant operators or quality engineers.

In light of this limitation of SPC methods, considerable research efforts have been expended on developing methodologies for root-cause identification. The basic approach among the reported methodologies is to use a diagnostic fault-quality model, which connects the measured product quality characteristic to process faults. For example, in a machining operation, dimensional features such as the position, orientation, and size of a machined feature are affected by deviations in process variables such as fixturing errors and/or machine tool errors. In this example, the dimensional features are the product quality characteristics and are treated as the outputs of a fault-quality model, while the errors in the process variables are the process faults and thus are the inputs to the model.

A wide variety of approaches can lead to the development of a diagnostic fault-quality model. Ceglarek and Shi (1996) linked process faults and the product quality of assembly processes using principal components. A linear fault-quality model of an explicit input-output format was

* Corresponding author.

presented for assembly processes by Apley and Shi (1998). Their work was extended to multistage assembly processes by Jin and Shi (1999) and Ding *et al*. (2002a). Djurdjanovic and Ni (2001) and Zhou *et al*. (2003b) developed a linear-deviation propagation model for multistage machining processes.

All of the above models were developed on the basis of governing physical laws in the corresponding processes. They bear the same linear-model structure because the process faults are assumed to be of a smaller magnitude as compared to nominal product dimensions. A linear fault-quality model can be generally expressed as:

$$\mathbf{y} = \mathbf{A}\mathbf{f} + \varepsilon, \tag{1}$$

where $\mathbf{y}$ is a $n \times 1$ vector of product quality measurements, $\mathbf{A}$ is a $n \times p$ constant system matrix determined by process/product designs, $\mathbf{f}$ is a $p \times 1$ random vector representing the process faults, and ε is a $n \times 1$ random vector representing the influence of measurement noise, un-modeled faults, and high-order nonlinear terms. Without loss of generality, we can further assume that the columns of the $\mathbf{A}$ matrix are of unit length. This can be achieved through a simple scaling process. In this paper, we focus on variation-source identification. Therefore, the process fault vector $\mathbf{f}$ is modeled as a random vector to describe the variation errors, as opposed to a mean-shift type of error in the process. Since the mean of quality measurements can always be subtracted, it is also assumed that the means of $\mathbf{f}$ and ε are zero.

After the fault-quality model is established, either statistical estimation methods or pattern matching methods can be used to identify which process fault has occurred based on product quality measurements.

In estimation methods, the fault-quality model of equation (1) is treated as a linear mixed model. The variances of the process faults ($\sigma^2_{f_1}, \ldots, \sigma^2_{f_p}$) are the variance components to be estimated in this mixed model (Searle *et al*., 1992; McCulloch and Searle, 2001). Based on the mixed model, Apley and Shi (1998) and Chang and Gossard (1998) used ordinary least-squares to estimate the random input $\hat{\mathbf{f}}$ and then calculate its variance as if $\hat{\mathbf{f}}$ was directly measured. Zhou *et al*. (2004) used a maximum-likelihood estimator and also provided the confidence intervals of the estimated variance of $\mathbf{f}$. Ding *et al*. (2005) compared different variance estimation methods and provided guidelines on method selection under different circumstances. Furthermore, Ding *et al*. (2002b) and Zhou *et al*. (2003a) performed diagnosability studies to identify the necessary conditions that need to be satisfied for an estimation method to be applicable.

Although based on the same fault-quality model, the pattern matching method is different from statistical estimation methods. The basic idea of the pattern matching technique is illustrated in Fig. 1. First, based on the fault-quality model, we can obtain the signatures of potential faults. Then the pattern of an occurring fault is extracted from measured data. Finally, the occurring fault can be identified if there is a match between the patterns of the

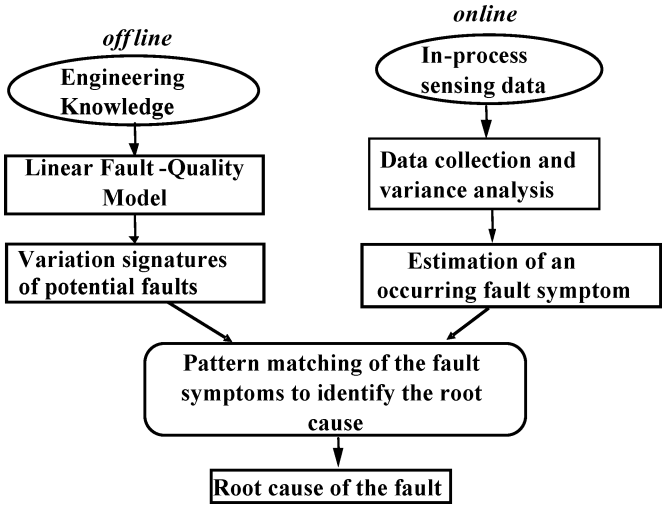

Fig. 1. Outline of the pattern matching method.

fault symptom and the fault signature. In Equation (1), the column vectors of $\mathbf{A}$ determine how a specific process fault affects the product quality characteristics and thus they are the *fault signature vectors*. Principal Component Analysis (PCA) (Johnson and Wichern, 1998) is the major statistical tool used to extract the *fault symptom* from measured data in the pattern matching method.

Ceglarek and Shi (1996) developed a pattern matching method for fixture fault diagnosis in automotive body assembly processes. The fault signatures were defined based on design information on the assembled part under faulty fixture conditions. After that, the pattern matching technique was used to map the extracted fault symptoms from measured data through PCA to the predefined fault signatures. The authors assumed a structured noise (the covariance matrix of ε, $\mathbf{\Sigma}_\varepsilon$, is diagonal) and a large sample size (namely, the covariance matrix of measurements $\mathbf{y}$, $\mathbf{\Sigma}_\mathbf{y}$, is known). This method was extended by Rong *et al*. (2000) to the compliant beam structure assembly model by considering the sample properties of the principal eigenvector of $\mathbf{\Sigma}_\mathbf{y}$ but keeping the structured noise requirement. Ding *et al*. (2002a) studied the impact of an unstructured noise under the large sample condition. They used a state space model for multistage manufacturing processes and generated the fault signatures based on the model. A PCA-based methodology was adopted to identify the single variation source in the process.

The statistical estimation method and the pattern matching method have different strengths. From a practicability point of view, the pattern matching method is very intuitive and possesses a clear geometric interpretation, which may help practitioners to understand and eliminate the variation source. Thus, the pattern matching method is easy for practitioners to implement and execute as compared with statistical estimation methods. On the other hand, the statistical estimation method allows people to analyze the test performance analytically because the testing statistics are tractable. The available statistical estimation methods for variation-source identification assume that the covariance

matrix of the noise term ε is in the simple form $\sigma_\varepsilon^2\mathbf{I}$. Unstructured noise is not considered.

This paper focuses on an extension of the pattern matching technique by considering both unstructured noise and sample uncertainty in the matching. The proposed method is robust and maintains a good identification probability. Using a machining process as an illustrative example, the paper demonstrates that current pattern matching procedures can have a remarkably low identification probability when the assumptions are not strictly satisfied. By contrast, our proposed method is more robust, maintaining a much higher identification probability, and is a preferable tool for variation-source identification in manufacturing quality improvement.

This paper is organized as follows. In Section 2, we present a robust pattern matching procedure for fault diagnosis. Section 3 uses a case study to illustrate the effectiveness of the proposed technique. Finally, we conclude the paper in Section 4.

2. Pattern matching techniques for root-cause identification

2.1. *Problem formulation and assumptions*

The linear fault-quality model as defined in Equation (1) is adopted in this paper. Assuming that the process fault $\mathbf{f}$ is independent of system noise ε, we can obtain the following from Equation (1):

$$\mathbf{\Sigma_y} = \mathbf{A\Sigma_f A}^T + \mathbf{\Sigma}_\varepsilon, \tag{2}$$

where $\mathbf{\Sigma_f}$ and $\mathbf{\Sigma}_\varepsilon$ are the covariance matrices of $\mathbf{f}$ and ε, respectively. Furthermore, it is usually assumed that process faults (i.e., the elements in $\mathbf{f}$) are independent of one another, and hence, $\mathbf{\Sigma_f}$ is a diagonal matrix.

Without loss of generality, let us assume that the ith fault occurs and only a single diagonal element of $\mathbf{\Sigma_f}$, σ_i^2, is nonzero. Then, we have:

$$\mathbf{\Sigma_y} = \sigma_i^2\mathbf{a}_i\mathbf{a}_i^T + \mathbf{\Sigma}_\varepsilon, \tag{3}$$

where $\mathbf{a}_i$ is the ith column vector of $\mathbf{A}$. If we multiply both sides of Equation (3) by $\mathbf{a}_i$, we have $\mathbf{\Sigma_y a}_i = \sigma_i^2(\mathbf{a}_i^T\mathbf{a}_i)\mathbf{a}_i + \mathbf{\Sigma}_\varepsilon\mathbf{a}_i$. If the components in ε are independent of one another and have the same variances, then $\mathbf{\Sigma}_\varepsilon$ will have the simple form $\sigma_\varepsilon^2\mathbf{I}$, where σ_ε^2 is the variance of the noise and $\mathbf{I}$ is an identity matrix with appropriate dimension. Under these conditions, we have:

$$\mathbf{\Sigma_y a}_i = (\sigma_i^2(\mathbf{a}_i^T\mathbf{a}_i) + \sigma_\varepsilon^2)\mathbf{a}_i. \tag{4}$$

Since $\sigma_i^2(\mathbf{a}_i^T\mathbf{a}_i) + \sigma_\varepsilon^2$ is a scalar, $\mathbf{a}_i$ is apparently an eigenvector of $\mathbf{\Sigma_y}$.

The above analysis indicates that when the ith process fault occurs, the eigenvector associated with the largest eigenvalue of $\mathbf{\Sigma_y}$, known as its *principal eigenvector*, should match the ith column vector of $\mathbf{A}$. Thus, the principal eigenvector of $\mathbf{\Sigma_y}$ is the *fault symptom* and will be used for root-cause identification in the pattern matching method.

Theoretically, fault signature $\mathbf{a}_i$ will exactly match the principal eigenvector of $\mathbf{\Sigma_y}$ with a noise term where $\mathbf{\Sigma}_\varepsilon = \sigma_\varepsilon^2\mathbf{I}$. This type of noise term is called a *structured noise*, referring to its special covariance matrix structure. In reality, not every noise term can be assumed to have such a simple covariance structure. In the presence of unstructured noise, where $\mathbf{\Sigma}_\varepsilon$ is not in the form of $\sigma_\varepsilon^2\mathbf{I}$, the direction of a fault symptom will deviate from the corresponding fault signature. Another difficulty results from the fact that the population covariance matrix $\mathbf{\Sigma_y}$ is never exactly known and will have to be estimated from the sample covariance matrix $\mathbf{S}_y$ of the measured data. When sampling from data, the principal eigenvector of $\mathbf{S}_y$ is a random vector, meaning that we will only know the sample of a fault symptom instead of its exact value.

As stated in the Introduction, the influences of these two effects on pattern matching procedures have previously been studied separately, by assuming either the absence of unstructured noise or the availability of large samples. Since these two assumptions cannot always hold in actual manufacturing processes, we will develop a robust pattern matching procedure by considering both the unstructured noise effect and the sampling effect together in this paper. We list our assumptions as follows:

1. The fault-quality relation can be adequately described by Equation (1) and the constant matrix $\mathbf{A}$ is known. The process faults and the system noises are assumed to be independent of one another.
2. The covariance of ε is in a general form as $\mathbf{\Sigma}_\varepsilon$. $\mathbf{\Sigma}_\varepsilon$ may be unknown but the range of its eigenvalues is assumed known.
3. One fault occurs at a time, meaning that no multiple faults occur in the system simultaneously.

The rationale of assumption (2) is that ε is usually dominated by measurement noise, thus the range of the eigenvalues represents the range of sensor accuracy that can be obtained through the sensor vendor's specification. Even if ε comprises unmodeled faults and high-order nonlinear residuals, an offline calibration could provide information on the eigenvalue range of $\mathbf{\Sigma}_\varepsilon$. Assumption (3) is not very restrictive because the probability of simultaneous multiple faults is small in many cases. Moreover, whenever a fault occurs, we need to identify and eliminate the fault rather than wait until another fault occurs.

2.2. *Robust Pattern Matching Technique*

As mentioned in Section 2.1, it is unrealistic to assume that $\mathbf{\Sigma}_\varepsilon$ is in the form of $\sigma_\varepsilon^2\mathbf{I}$, which can probably only closely represent the case where identical measurement devices are used to measure quality features. When different measurement devices are used, the variances of the different components in ε will be different. Moreover, the noise

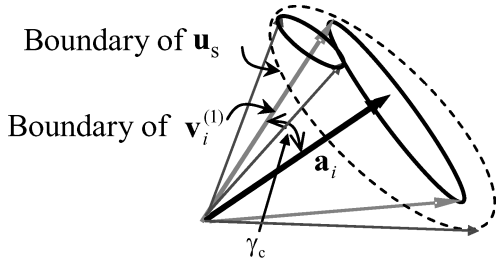

Fig. 2. The boundary for pattern matching of the columns of **A**.

term ε could comprise un-modeled faults and high-order nonlinear residuals, which are more likely to be correlated and makes a diagonal structure in $\mathbf{\Sigma}_\varepsilon$ even less likely. For this reason, it is more practical to study $\mathbf{\Sigma}_\varepsilon$ in a general structure.

In addition to disturbances due to unstructured noises, sampling uncertainty in $\mathbf{S_y}$ makes the sample principal eigenvector a random vector. Instead of checking if the sample principal eigenvector is equal to the theoretical fault signature in an exact and deterministic sense, we should establish a confidence boundary to see if they are equal in a statistical sense.

Figure 2 illustrates the joint effects of the aforementioned two kinds of uncertainty. Vector $\mathbf{a}_i$ is a column vector of **A**, which is the same as the principal eigenvector of $\mathbf{A\Sigma_f A}^T$ when the ith fault occurs. With the presence of an unstructured noise, the principal eigenvector of $\mathbf{\Sigma_y}$, denoted as $\mathbf{v}_i^{(1)}$ (the subscript i means the ith fault occurs, while the superscript$^{(1)}$ represents that it is the principal eigenvector), is not the same as $\mathbf{a}_i$. Instead, it will fall in a cone as shown in Fig. 2. The boundary of the cone can be represented by the angle γ_c. The sample principal eigenvector of $\mathbf{S_y}$, denoted as $\mathbf{u}_s$, will be different from $\mathbf{v}_i^{(1)}$ due to sample uncertainty. The final boundary for $\mathbf{u}_s$ can be illustrated by the dashed-line as shown in Fig. 2. If the sample principal eigenvector falls in the confidence boundary of a fault signature vector $\mathbf{a}_i$, we can claim that the corresponding fault occurs. Note that Fig. 2 exaggerates the size of the region for the sake of illustration. When the sample size is large and the noise variance is small, the confidence range of $\mathbf{u}_s$ is actually small.

2.2.1. *Disturbance due to an unstructured noise*

In this part we focus on the perturbation of an unstructured noise on the eigenvectors of $\mathbf{\Sigma_y}$. When the ith fault occurs, the *principal* eigenvalue and eigenvector pair of $\mathbf{A\Sigma_f A}^T$, $\mathbf{\Sigma_y}$, and $\mathbf{S_y}$ are $\{\lambda_{\mathbf{f},i}^{(1)}, \mathbf{a}_i\}$, $\{\lambda_{\mathbf{y},i}^{(1)}, \mathbf{v}_i^{(1)}\}$ and $\{l_{\mathbf{y},i}^{(1)}, \mathbf{u}_s\}$, respectively. With the presence of an unstructured noise the principal eigenvector ($\mathbf{v}_i^{(1)}$) of $\mathbf{\Sigma_y}$ differs from $\mathbf{a}_i$ and the boundary of this difference is related to the additive noise covariance matrix $\mathbf{\Sigma}_\varepsilon$.

A useful result regarding the perturbation boundary caused by an unstructured noise is given by Ding *et al.* (2002a): when only the ith fault occurs in the system (i.e., only $\lambda_{\mathbf{f},i}^{(1)}$, the *principal* eigenvalue of $\mathbf{A\Sigma_f A}^T$ is nonzero) and $\|\mathbf{\Sigma}_\varepsilon\|_2 \le \lambda_{\mathbf{f},i}^{(1)}/4$, then:

$$\text{dist}(\text{span}\{\mathbf{a}_i\}, \text{span}\{\mathbf{v}_i^{(1)}\}) \le \left(\frac{4}{\lambda_{\mathbf{f},i}^{(1)}}\sqrt{\lambda_{\max}^2(\mathbf{\Sigma}_\varepsilon) - \lambda_{\min}^2(\mathbf{\Sigma}_\varepsilon)}\right), \tag{5}$$

where $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)$ and $\lambda_{\min}(\mathbf{\Sigma}_\varepsilon)$ are the largest and the smallest eigenvalues of $\mathbf{\Sigma}_\varepsilon$, and dist(span$\{\mathbf{a}_i\}$, span$\{\mathbf{v}_i^{(1)}\}$) is the *distance* (Golub and Van Loan, 1996) between the space spanned by $\mathbf{a}_i$ and $\mathbf{v}_i^{(1)}$ respectively. The condition $\|\mathbf{\Sigma}_\varepsilon\|_2 \le \lambda_{\mathbf{f},i}^{(1)}/4$ means that the largest variance of the noise term is four times smaller than the variance of the process fault, which is not restrictive in practice.

Since $\mathbf{a}_i$, $\mathbf{v}_i^{(1)}$ are two unit vectors (in the sense of its 2-norm) in R^n, the distance between two subspaces of span$\{\mathbf{a}_i\}$, span$\{\mathbf{v}_i^{(1)}\}$ is equal to the sine of the angle ($\Delta\theta$) between $\mathbf{a}_i$ and $\mathbf{v}_i^{(1)}$. That is, $\sin(\Delta\theta) = \text{dist}(\text{span}\{\mathbf{a}_i\}, \text{span}\{\mathbf{v}_i^{(1)}\})$ (The proof of this result can be found in the Appendix). Now we have the following result:

$$\Delta\theta \le \sin^{-1}\left(\frac{4}{\lambda_{\mathbf{f},i}^{(1)}}\sqrt{\lambda_{\max}^2(\mathbf{\Sigma}_\varepsilon) - \lambda_{\min}^2(\mathbf{\Sigma}_\varepsilon)}\right) = \gamma_c. \tag{6}$$

Obviously the angle between $\mathbf{a}_i$ and $\mathbf{v}_i^{(1)}$ can be calculated by using $\cos(\Delta\theta) = \mathbf{v}_i^{(1)T}\mathbf{a}_i$. Its boundary will be denoted by:

$$\gamma_c = \sin^{-1}\left(\frac{4}{\lambda_{\mathbf{f},i}^{(1)}}\sqrt{\lambda_{\max}^2(\mathbf{\Sigma}_\varepsilon) - \lambda_{\min}^2(\mathbf{\Sigma}_\varepsilon)}\right),$$

hereafter. The above result indicates that the difference between $\mathbf{a}_i$ and $\mathbf{v}_i^{(1)}$, represented by the angle $\Delta\theta$ between them (as shown in Fig. 2), is determined by the eigenvalue of $\mathbf{A\Sigma_f A}^T$ and the extreme eigenvalues of ($\mathbf{\Sigma}_\varepsilon$). We would also like to point out that the boundary specified by Equation (6) is a worst-case boundary.

Inequality (6) suggests that the perturbation boundary between $\mathbf{a}_i$ and $\mathbf{v}_i^{(1)}$ depends on $\lambda_{\mathbf{f},i}^{(1)}/\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)$ and $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)/\lambda_{\min}(\mathbf{\Sigma}_\varepsilon)$. The value of $\lambda_{\mathbf{f},i}^{(1)}/\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)$ can be somewhat regarded as the signal-to-noise ratio, while $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)/\lambda_{\min}(\mathbf{\Sigma}_\varepsilon)$ indicates the imbalance in accuracy associated with different measurement devices. With an increase in signal-to-noise ratio, the perturbation boundary will get smaller. For instance, if $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon) = 0.030$, $\lambda_{\mathbf{f},i}^{(1)}/\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)$ is 10, and $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)/\lambda_{\min}(\mathbf{\Sigma}_\varepsilon)$ is around ten, the perturbation boundary (i.e., the angle) will be around 20°. However, if $\lambda_{\mathbf{f},i}^{(1)}/\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)$ is around 50, meaning the fault magnitude is five times larger than before, the perturbation angle will reduce to 5° or so. On the other hand, with a higher imbalance among eigenvalues of $\mathbf{\Sigma}_\varepsilon$, the perturbation boundary will become larger. In the case that all eigenvalues of $\mathbf{\Sigma}_\varepsilon$ are the same, there will be no perturbation on $\mathbf{v}_i^{(1)}$.

2.2.2. *Disturbance due to sampling uncertainty*

This section presents a testing procedure for the principal eigenvector of a sample covariance matrix. We use the same notation as before that $\mathbf{u}_s$ and $\mathbf{v}_i^{(1)}$ are the principal eigenvector of $\mathbf{S}_y$ and $\mathbf{\Sigma}_\mathbf{y}$, respectively, when the ith fault occurs in the system. According to Murihead (1982), if $\lambda_{\mathbf{y}i}^1$ is a distinct eigenvalue (which is true in our case when a single fault occurs) and $\mathbf{y}$ follows a normal distribution, then $(N\text{-}1)^{1/2}(\mathbf{u}_s - \mathbf{v}_i^{(1)})$ asymptotically follows a n-variate normal distribution $N(\mathbf{0}, \mathbf{\Gamma})$, where:

$$\mathbf{\Gamma} = \lambda_{\mathbf{y},i}^{(1)} \sum_{j=2}^{n} \frac{\lambda_{\mathbf{y},i}^{(j)}}{\left(\lambda_{\mathbf{y},i}^{(1)} - \lambda_{\mathbf{y},i}^{(j)}\right)^2} \mathbf{v}_i^{(j)} \mathbf{v}_i^{(j)T}, \tag{7}$$

N is the sample size, and n is the dimension of measurement in Equation (1). Apparently, due to the sampling uncertainty, different realizations of $\mathbf{u}_s$ will generate a hyper-dimensional ellipsoid centered around the corresponding vector $\mathbf{v}_i^{(1)}$. Following the above result, we know the limiting distribution of $W_i \equiv (N-1)(\mathbf{v}_i^{(1)} - \mathbf{u}_s)^T (l_{\mathbf{y},i}^{(1)} \mathbf{S}_\mathbf{y}^{-1} - 2\mathbf{I} + 1/l_{\mathbf{y},i}^{(1)} \mathbf{S}_\mathbf{y})(\mathbf{v}_i^{(1)} - \mathbf{u}_s)$ is χ_{n-1}^2 (Murihead, 1982).

Given one fault occurring in the system, from the measurements $\mathbf{y}$ we can get the fault symptom vector $\mathbf{u}_s$, which is the principal eigenvector of $\mathbf{S}_y$. The objective of variation-source identification is to find out which one of p potential faults, represented by the p column vectors of $\mathbf{A}$, occurred in the system based on the value of $\mathbf{u}_s$. To reach this goal, a series of asymptotic hypothesis tests can be formulated as follows. If we denote $\mathbf{v}_s$ as the principal eigenvector of the population covariance matrix of the current quality measurements and $\mathbf{v}_k^{(1)}$ as the principal eigenvector of $\mathbf{\Sigma}_\mathbf{y}$ when the kth fault occurs in the process, then for $k = 1, 2, \ldots, p$, we can formulate the test as:

$$H_0 : \mathbf{v}_s = \mathbf{v}_k^{(1)},$$
$$H_1 : \mathbf{v}_s \neq \mathbf{v}_k^{(1)}.$$

The testing statistic can be developed as follows based on Equation (7):

$$W_k \equiv (N-1)\left(\mathbf{v}_k^{(1)} - \mathbf{u}_s\right)^T \left(l_{\mathbf{y},k}^{(1)} \mathbf{S}_\mathbf{y}^{-1} - 2\mathbf{I} + \frac{1}{l_{\mathbf{y},k}^{(1)}} \mathbf{S}_\mathbf{y}\right) \left(\mathbf{v}_k^{(1)} - \mathbf{u}_s\right). \tag{8}$$

As indicated above, it is known that W_k asymptotically follows the distribution of χ_{n-1}^2 under the null hypothesis. If $W_k > \chi_{n-1}^2(\alpha)$, where $\chi_{n-1}^2(\alpha)$ is the upper $100\alpha\%$ point of the χ_{n-1}^2 distribution, we reject the null hypothesis, meaning that $\mathbf{v}_k^{(1)}$ is not the principal eigenvector of the current $\mathbf{\Sigma}_y$. Otherwise, we accept the null hypothesis, meaning that $\mathbf{v}_k^{(1)}$ is the principal eigenvector of the current $\mathbf{\Sigma}_y$, and thus, we know that the kth fault occurred in the system.

A difficulty in conducting this test is that the values of $\mathbf{v}_k^{(1)}$, $k = 1, 2, \ldots, p$, are unknown. A robust matching procedure is developed in the next section to deal with this difficulty.

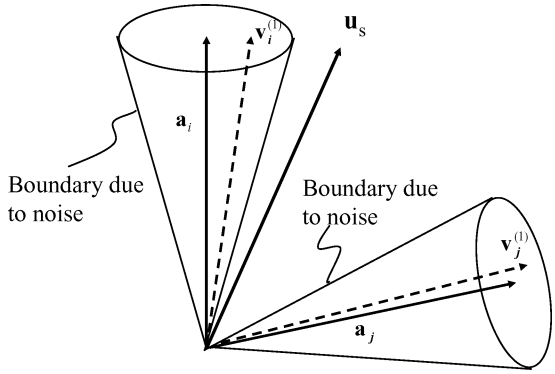

Fig. 3. The method to identify the boundary of $\mathbf{v}_i^{(1)}$.

2.2.3. *A robust pattern matching procedure*

A robust pattern matching procedure that considers the effects of noise perturbation and also sampling uncertainty is developed in this section. Consider two faults, represented by $\mathbf{a}_i$ and $\mathbf{a}_j$, in Fig. 3. Vectors $\mathbf{a}_i$ and $\mathbf{a}_j$ are the fault signature vectors from matrix $\mathbf{A}$. Due to the perturbation of unstructured noises, there are two cones centered around $\mathbf{a}_i$ and $\mathbf{a}_j$, respectively, within which the eigenvectors of $\mathbf{\Sigma}_\mathbf{y}$, $\mathbf{v}_i^{(1)}$ and $\mathbf{v}_j^{(1)}$, will fall. Once measurement data are collected, the fault symptom vector $\mathbf{u}_s$, the principal eigenvector, can be calculated from $\mathbf{S}_y$.

We want to identify which fault occurs exactly. There are two cases: (i) when $\mathbf{u}_s$ falls into either one of the two cones; and (ii) when $\mathbf{u}_s$ falls outside both cones. For the first, because $\mathbf{u}_s$ falls within the confidence boundary of the population covariance matrices, it means that the corresponding fault occurs. The decision rule is then simple and is the same as if there is no sampling uncertainty, namely, we can immediately claim that a fault represented by either $\mathbf{a}_i$ or $\mathbf{a}_j$, whichever is applicable, occurs. For the second case, we need to further calculate a test statistic similar to that in Equation (8) to decide which fault has occurred.

However, due to the perturbation from unstructured noise ε, we do not know the precise position of $\mathbf{v}_k^{(1)}$, $k = 1, 2, \ldots, p$. Instead we only know the worst-case boundary, within which $\mathbf{v}_k^{(1)}$ falls. One straightforward approach would be using the worst-case boundary as the position of $\mathbf{v}_k^{(1)}$. However, a simple numerical study shows that the test statistic in Equation (8) using a worst case $\mathbf{v}_k^{(1)}$ (i.e., using the value on the worst-case boundary) often yields a very large W_k value so that the null hypothesis that $\mathbf{v}_k^{(1)}$ is the principal eigenvector of $\mathbf{\Sigma}_\mathbf{y}$ is always rejected and a high misidentification rate results.

In order to reduce the misidentification rate, we adopt a different approach to locate $\mathbf{v}_k^{(1)}$ for testing Equation (8). Instead of using a boundary vector, we try to find a vector located within the hyper-dimensional cone (including the boundary) that can minimize the W_k value. Then this vector is substituted into Equation (8) as $\mathbf{v}_k^{(1)}$ to calculate

W_k^*. A typical nonlinear programming problem to find an appropriate $\mathbf{v}_k^{(1)}$ can be formulated as follows:

$$\text{Objective: } W_k^* \equiv \min_{\mathbf{v}_k^{(1)}} (N-1)\left(\mathbf{v}_k^{(1)} - \mathbf{u}_s\right)^T \left(l_{\mathbf{y},k}^{(1)} \mathbf{S}_\mathbf{y}^{-1} - 2\mathbf{I} + \frac{1}{l_{\mathbf{y},k}^{(1)}} \mathbf{S}_\mathbf{y}\right)\left(\mathbf{v}_k^{(1)} - \mathbf{u}_s\right),$$

$$\text{Constraints: } \mathbf{v}_k^{(1)T}\mathbf{v}_k^{(1)} = 1$$

$$\Delta\theta_k = \cos^{-1}\left(\mathbf{v}_k^{(1)T}\mathbf{a}_k\right) \le \gamma_c$$

In the above formulation, the objective function is to find a vector $\mathbf{v}_k^{(1)}$ that minimizes W_k, while satisfying a set of constraints. The constraints mean that $\mathbf{v}_k^{(1)}$ should remain as a unit vector and the angle between $\mathbf{v}_k^{(1)}$ and $\mathbf{a}_k$ should be smaller than the worst-case boundary range γ_c.

Based on the Kuhn-Tucker conditions (Taha, 1992), we can convert the inequality constraint into an equality constraint and calculate the solution of $\mathbf{v}_k^{(1)}$. In our numerical study, we use the `fmincon` function in the Matlab software to solve this nonlinear programming problem. After finding the optimal $\mathbf{v}_k^{(1)}$, the final step is to calculate the value of W_k^* and compare this value with the upper $100\alpha\%$ point of the χ^2_{n-1} distribution.

Our decision rule using the above optimized $\mathbf{v}_k^{(1)}$ and W_k^* follows what is previously stated: if the W_k^* is not greater than the critical value, we claim that the corresponding fault occurs, otherwise we claim this fault does not occur. This conclusion is conservative in terms of fault identification. Because the vector that minimizes the W_k^* statistic might not be the true $\mathbf{v}_k^{(1)}$, we are more inclined to conclude that a fault is indeed occurring. Subsequent investigations are still needed to verify the conclusion. From our experiences, we find that plant engineers and management welcome this conservation. They would prefer a small misidentification rate and are very cautious about any potential signs of significant process faults because of the very high warranty cost associated with quality problems.

Under some conditions, for example, when the angle between two signature vectors $\mathbf{a}_i$ and $\mathbf{a}_j$ is quite small, our decision rule may lead to the conclusion that both of these two faults occurred. This problem of "fault isolation," i.e., how to distinguish among two or more faults, is of considerable interest and the isolation condition under which all faults are uniquely identifiable worth further investigation.

After identifying the ith process fault, we can further estimate the variance magnitude of the fault. We have (Golub and Van Loan, 1996):

$$\lambda_{\mathbf{f},i}^{(1)} + \lambda_{\min}(\mathbf{\Sigma}_\varepsilon) \le \lambda_{\mathbf{y},i}^{(1)} \le \lambda_{\mathbf{f},i}^{(1)} + \lambda_{\max}(\mathbf{\Sigma}_\varepsilon). \tag{9}$$

Based on Equation (9), we can get the range of eigenvalue of $\mathbf{A}\mathbf{\Sigma}_\mathbf{f}\mathbf{A}^T$ as follows:

$$\lambda_{\mathbf{y},i}^{(1)} - \lambda_{\max}(\mathbf{\Sigma}_\varepsilon) \le \lambda_{\mathbf{f},i}^{(1)} \le \lambda_{\mathbf{y},i}^{(1)} - \lambda_{\min}(\mathbf{\Sigma}_\varepsilon). \tag{10}$$

From Equation (4), we know that $\sigma_i^2(\mathbf{a}_i^T\mathbf{a}_i)$ is the eigenvalue of $\mathbf{A}\mathbf{\Sigma}_\mathbf{f}\mathbf{A}^T$ corresponding to the process fault. Substituting $\lambda_{\mathbf{f},i} = \sigma_i^2(\mathbf{a}_i^T\mathbf{a}_i)$ into Equation (10) and using the average of the upper and lower bounds in Equation (10) as an approximation for σ_i^2, we then estimate the variation magnitude of a fault as:

$$\sigma_i^2 \approx \frac{1}{2(\mathbf{a}_i^T\mathbf{a}_i)}\left\{\left[l_{\mathbf{y},i}^{(1)} - \lambda_{\min}(\mathbf{\Sigma}_\varepsilon)\right] + \left[l_{\mathbf{y},i}^{(1)} - \lambda_{\max}(\mathbf{\Sigma}_\varepsilon)\right]\right\}, \tag{11}$$

where $\lambda_{\mathbf{y},i}^{(1)}$ is substituted by its sample version $l_{\mathbf{y},i}^{(1)}$. Apparently, if the variance of the noise is relatively small as compared to that of a process fault, which is usually the case, the approximation error will be small. The reason for this is that when the largest and smallest eigenvalues of $\mathbf{\Sigma}_\mathbf{y}$ become smaller, so does the range stated in Equation (10).

To summarize, the proposed pattern matching procedure is listed as follows.

Step 1. A fault-quality model $\mathbf{y} = \mathbf{A}\mathbf{f} + \varepsilon$ is developed. Based on the model, fault signature vectors $\mathbf{a}_k$, $k = 1 \ldots, p$, are obtained from the columns of $\mathbf{A}$.

Step 2. The multivariate measurements $\mathbf{y}$ of the product quality features are obtained during production. The sample size is N and the dimension of the measurements is n.

Step 3. Based on $\mathbf{y}$, calculate the sample covariance matrix $\mathbf{S}_\mathbf{y}$ and its principal eigenvector $\mathbf{u}_s$ using PCA.

Step 4. Estimate $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)$ and $\lambda_{\min}(\mathbf{\Sigma}_\varepsilon)$ from an analysis of the accuracy specification of the measurement system. Based on Equation (6), we can calculate γ_c. Here we often need to substitute $\lambda_{\mathbf{f},i}^{(1)}$ with the tolerance specification of the corresponding process variable if the precise value of $\lambda_{\mathbf{f},i}^{(1)}$ is not yet known. Then, we calculate the angles $\Delta\theta_k$ between $\mathbf{a}_k$ and $\mathbf{u}_s$, $k = 1 \ldots, p$. If all of the angles are larger than γ_c, go to next step, otherwise go to Step 6.

Step 5. Select a significance level $\alpha \times 100\%$. For the nonlinear programming problem, we find the optimal value of $\mathbf{v}_k^{(1)}$, $k = 1 \ldots, p$, and calculate the value of $W_k^*, k = 1 \ldots, p$. If all the values are larger than $\chi^2_{n-1}(\alpha)$, we can claim that none of the known faults occurs, otherwise go to next step.

Step 6. If $\Delta\theta_k \le \gamma_c$ or $W_k^* \le \chi^2_{n-1}(\alpha)$, we can claim that the kth fault, represented by the kth column of matrix $\mathbf{A}$, occurs. After that, the variation magnitude of the fault can be calculated using Equation (11). If all $W_k^*, k = 1 \ldots, p$, are larger than $\chi^2_{n-1}(\alpha)$, then we claim that none of the known faults occur.

3. Case study

3.1. *Introduction to the experimental machining process*

The manufacturing process considered here is a three-stage machining process. The product is a V-6 automotive engine head. Its key features are shown in Fig. 4.

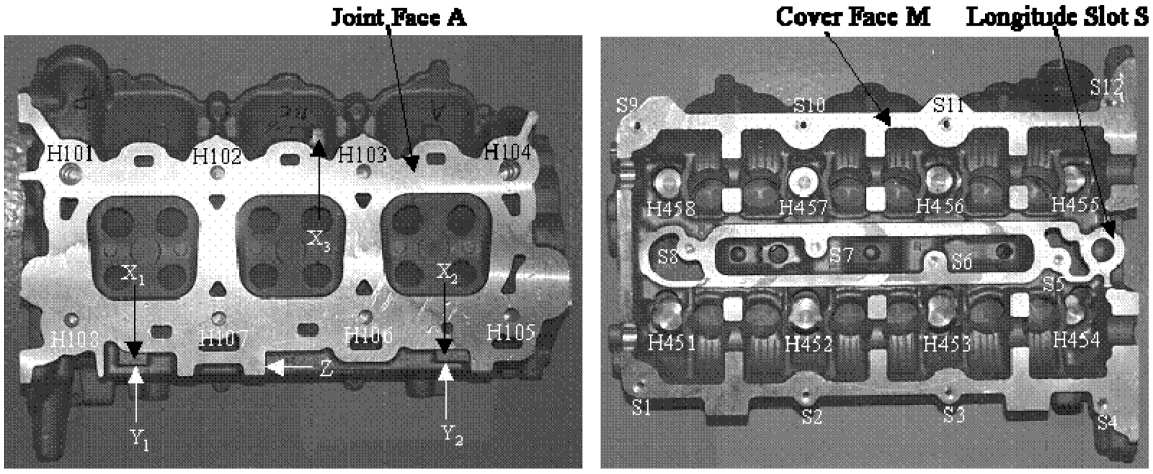

Fig. 4. The final product and its key features: (a) the joint face of the engine head; and (b) the cover face of the engine head.

The key features include the cover face, joint face, datum locating holes (H101 and H104 in Fig. 4), the slot on the cover face, and extra holes on the cover and joint face (H101–H108, H451–H458, S1–S12 in Fig. 4). Besides these features, X_1, X_2, X_3, Y_1, Y_2, Z are the initial locating points on the raw casting workpiece. They are used as datum for the initial cutting. Each operation and corresponding datum setup is shown in Fig. 5.

In this process, the key dimensions of the engine head are measured on a Coordinate Measurement Machine (CMM). A total of 15 (or 16) points on the joint (or cover) face are measured to determine the quality of the machining operation. These measurement points are evenly distributed on the two surfaces. Therefore, $\mathbf{y}$ will be a 31×1 vector (i.e., $n = 31$), consisting of the deviations at 15 points on the cover face and 16 points on the joint face.

Based on the CMM measurements of these features from the machining operations, potential process faults are identified as the locating pin position errors of the first and the second stages of the process. The variation source is then a 6×1 vector (i.e., $p = 6$), where the first three elements correspond to three pins at the first stage and the last three elements correspond to the three pins at the second stage.

A fault-quality model ($\mathbf{y} = \mathbf{A}_\mathrm{m}\mathbf{f}_\mathrm{m} + \varepsilon$) bearing the same linear structure as Equation (1) can be established from a sophisticated kinematic analysis (please refer to Zhou *et al.* (2003b) for details). The coefficient matrix $\mathbf{A}_\mathrm{m}$ of this specific process is given as:

$$\mathbf{A}_\mathrm{m} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}, \tag{12}$$

where $\mathbf{A}_1 = [\mathbf{A}_{11} \quad \mathbf{0}]$, $\mathbf{A}_2 = [\mathbf{0} \quad \mathbf{A}_{22}]$, and

$$\mathbf{A}_{11} = \begin{bmatrix} -0.5402 & -0.6455 & -0.2664 & 0.0683 & 0.5158 & 0.8555 & 0.9902 & 1.1863 & 1.3106 & 1.0177 & 0.6394 & 0.3403 & 0.0530 & -0.0820 & -0.2776 \\ 0.9690 & 0.5991 & 0.2203 & -0.1140 & -0.5612 & -0.9005 & -0.7058 & -0.4225 & -0.2428 & 0.1052 & 0.4597 & 0.7585 & 1.0455 & 0.8192 & 0.5367 \\ 0.5711 & 1.0464 & 1.0461 & 1.0458 & 1.0453 & 1.0450 & 0.7156 & 0.2362 & -0.0678 & -0.1229 & -0.0991 & -0.0988 & -0.0985 & 0.2628 & 0.7409 \end{bmatrix}^T,$$

$$\mathbf{A}_{22} = \begin{bmatrix} -0.6349 & -0.4590 & -0.0332 & 0.2510 & 0.6955 & 1.0215 & 1.1372 & 1.1155 & 0.9048 & 0.6472 & 0.5864 & 0.3875 & 0.0777 & -0.1490 & -0.4063 & 0.0776 \\ 0.9602 & 0.6817 & 0.2556 & -0.0288 & -0.4736 & -0.5857 & -0.5031 & -0.3477 & -0.1368 & 0.1210 & -0.0142 & 0.3998 & 0.7098 & 0.9368 & 0.9339 & 0.4496 \\ 0.6747 & 0.7773 & 0.7776 & 0.7778 & 0.7781 & 0.5641 & 0.3659 & 0.2322 & 0.2320 & 0.2318 & 0.4279 & 0.2127 & 0.2124 & 0.2123 & 0.4725 & 0.4728 \end{bmatrix}^T.$$

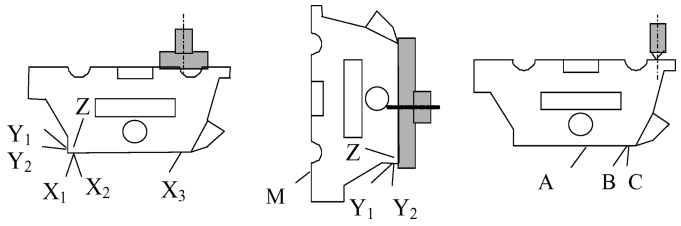

Fig. 5. The operation sequence.

Please note that the lengths of the columns in the coefficient matrix are not one in Equation (12). However, a simple scaling yields $\mathbf{y} = \mathbf{A}_\mathrm{m}\mathbf{f}_\mathrm{m} + \varepsilon = (\mathbf{A}_\mathrm{m}\mathbf{L}^{-1})(\mathbf{L}\mathbf{f}_\mathrm{m}) + \varepsilon = \mathbf{A}\mathbf{f} + \varepsilon$, where $\mathbf{A} = \mathbf{A}_\mathrm{m}\mathbf{L}^{-1}$, $\mathbf{f} = \mathbf{L}\mathbf{f}_\mathrm{m}$, $\mathbf{L} = diag(L_1, L_2, L_3, L_4, L_5, L_6)$, and L_i is the length of the ith column of matrix $\mathbf{A}_\mathrm{m}$ ($i = 1, 2, \ldots, 6$), and thus the length of each column of $\mathbf{A}$ is one. In this way, we can treat $\mathbf{L}\mathbf{f}_\mathrm{m}$ as the standardized variation source $\mathbf{f}$. The six column vectors of $\mathbf{A}$ are the signature vectors of the six potential faults.

In this example, the noise ε is dominated by errors associated with the measurement device. Since the orientations of a CMM probe are different when it is used to measure either

the joint face or the cover face, the variances of the measurement errors are therefore different for the measurement points on these two different surfaces. Based on engineering analysis, the standard deviations of components in ε are between 0.008 and 0.015 mm.

Furthermore, from the design specification of our fixture system, the locating tolerance is around 0.050 mm. If the standard deviation of components in **f** is larger than this tolerance, we deem that the corresponding fault occurs. Next we will perform root-cause identification using the pattern matching technique based on this fault-quality model and the CMM measurements.

3.2. *Pattern matching for root-cause identification in the machining process*

3.2.1. *Comparison of the developed technique and previously available techniques*

Given that the standard deviations of the components in ε are between 0.008 and 0.015 mm, we randomly generate a positive-definite matrix with diagonal elements falling in the range from $64 - 225 \times 10^{-6}$ mm^2. Then, we use this matrix as the noise covariance matrix, where the largest and smallest eigenvalues, $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)$ and $\lambda_{\min}(\mathbf{\Sigma}_\varepsilon)$, can be easily computed. We consider the situation that the third fault has occurred. The variations of **f** are specified accordingly as:

$$\text{var } \mathbf{f}(k) = \begin{cases} c_1 \text{ mm}^2, & k = 3, \\ c_2 \text{ mm}^2, & k \neq 3, \end{cases}$$

where c_1 is larger than 2.5×10^{-3}, the square of the locating tolerance, and c_2 is smaller than this value. The theoretical chi-square critical value of Equation (8) is $\chi^2_{30}(0.01) = 50.892$ and $\chi^2_{30}(0.05) = 43.773$ at significance levels of 0.01 and 0.05, respectively.

First, we would like to compare the misidentification rates of the proposed method with two prior methods developed by Ding *et al*. (2002a) and Rong *et al*. (2000) respectively. The former method considers the perturbation of unstructured noises but not the sampling uncertainty, whereas the latter considers sampling uncertainty in the absence of unstructured noises.

We conducted 1000 replicates of the fault identification using the aforementioned three procedures for a set of combinations of sample size N and perturbation angle γ_c. Table 1 lists the resulting means and standard deviations of the misidentification rates for different procedures at a 1% significance level. The misidentification rates are defined as the rate of the number of misidentification cases over the total number of tests (i.e., 1000 replications). We denote by e_1, e_2, and e_3 the misidentification rate obtained using the procedure of Ding *et al*. (2002a), Rong *et al*. (2000), and by this current paper, respectively. The mean and standard deviation values of the misidentification rate are calculated based on 40 iterations.

From the table we can observe the following.

1. The procedure of Ding *et al*. (2002a) results in very high misidentification rates when the sample size is small. For instance, the mean of e_1 is nearly 100% for a sample size as high as $N = 1000$ when $\gamma_c < 2.42^\circ$. The misidentification rate e_1 depends on the value of γ_c; given the same sample size, the larger the γ_c, the smaller the e_1. This is understandable because γ_c provides a worst-case boundary and a larger γ_c can help in offsetting disturbances from sampling uncertainty. When a smaller sample size is used such as $N = 50$, using the procedure of Ding *et al*. (2002a) we will still have a large misidentification rate even for a γ_c as large as 15.62°. Too large a γ_c is not preferable because it will allow the worst-case boundaries for different faults to overlap. On the other hand, if we have a large N (say $N = 1000$), the misidentification

Table 1. Misidentification rates of the three pattern matching methods

		γ_c *(degrees)*											
		0		*0.91*		*2.42*		*4.06*		*9.94*		*15.62*	
N	E	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*
50	e_1	100	(0.000)	100	(0.000)	100	(0.000)	96.49	(0.555)	66.81	(1.549)	45.67	(1.917)
	e_2	86.07	(0.976)	86.09	(0.768)	86.31	(0.907)	86.16	(0.819)	87.65	(0.788)	88.73	(0.764)
	e_3	81.43	(1.084)	33.64	(1.538)	10.02	(1.002)	0.46	(0.182)	0.03	(0.051)	0.01	(0.022)
100	e_1	100	(0.000)	100	(0.000)	99.95	(0.082)	76.19	(1.618)	29.66	(1.476)	14.87	(1.121)
	e_2	29.49	(1.633)	29.31	(1.617)	29.86	(1.522)	31.14	(1.477)	34.53	(1.562)	37.41	(1.662)
	e_3	22.15	(1.383)	1.07	(0.375)	0.17	(0.114)	0.00	(0.016)	0.00	(0.000)	0.00	(0.000)
400	e_1	100	(0.000)	99.89	(0.125)	69.92	(1.714)	7.19	(0.693)	0.25	(0.181)	0.03	(0.065)
	e_2	3.78	(0.658)	3.88	(0.637)	3.91	(0.615)	6.07	(0.734)	10.42	(0.952)	14.26	(1.180)
	e_3	2.67	(0.590)	0.01	(0.022)	0.00	(0.000)	0.00	(0.000)	0.00	(0.000)	0.00	(0.000)
1000	e_1	100	(0.000)	95.71	(0.583)	19.54	(1.090)	0.07	(0.091)	0.00	(0.000)	0.00	(0.000)
	e_2	2.64	(0.471)	2.72	(0.533)	2.79	(0.510)	8.01	(0.909)	21.25	(0.982)	33.19	(1.189)
	e_3	1.42	(0.308)	0.00	(0.000)	0.00	(0.000)	0.00	(0.000)	0.00	(0.000)	0.00	(0.000)

rate will indeed reduce to zero provided that there is a reasonable γ_c size (say $\gamma_c > 4.06^\circ$).

2. The procedure of Rong *et al*. (2000) results in a misidentification rate that increases very rapidly when γ_c is nonzero. This is particularly obvious for a large sample. The reason for this is that their method uses $\mathbf{a}_k$ instead of $\mathbf{v}_k^{(1)}$ in the statistic W_k in Equation (8). When γ_c increases, the deviation of $\mathbf{a}_k$ from $\mathbf{v}_k^{(1)}$ tends to increase and thus make the value of W_k larger. Therefore, we will be inclined to reject the hypothesis that the kth fault occurs. Given the same boundary γ_c, e_2 has a decreasing tendency when the sample size increases.
3. Using the procedure proposed in this paper, the misidentification rate is always smaller than the other two methods. In fact, except for the cases where $\gamma_c = 0$, the proposed method will have a near-zero misidentification rate. For $\gamma_c = 0$, using the procedure proposed in this paper will be equivalent to using the procedure in Rong *et al*. (2000) and e_3 is almost exactly the same as e_2 under that condition.

We can also find that the standard deviations of the misidentification rates of our method are similar to the other two methods and decrease consistently with increasing γ_c and N.

Usually, a low misidentification rate is accompanied by a high false alarm rate, which is the probability of identifying a normal working condition as a faulty working condition. Since we adopt a conservative approach to find W_i^* for diagnosis, it will not be surprising if the false alarm rate of the proposed method is higher than the other two methods. In this study, we used the same example as above but change the variances of the elements of $\mathbf{f}$ to be $c_1 = 20 \times 10^{-4}$ and $c_2 = 2 \times 10^{-4}$, both of which are smaller than the 2.5×10^{-3} tolerance level that corresponds to the case that no fault occurs. We again replicated our numerical study 1000 times and list the resulting means and standard deviations of the false alarm rates in Table 2, at a 1% significance level, where the false alarm rate δ is defined as the number of false alarms over the total number of tests (1000 replicates). Similarly to the notation used for misidentification rate, δ_1, δ_2, and δ_3 represent the fault alarm rates of the three investigated methods. As in Table 1, the mean and standard deviation values of the false alarm rates are calculated based on 40 iterations.

From this table we find that δ_3 is larger than δ_1 and δ_2 under almost all circumstances. As explained above, this is not surprising. However, the false alarm rate using our proposed procedure is not alarmingly high. When either there is an unstructured noise disturbance (say, $\gamma_c > 4.06^\circ$) or the sample size is relatively large (say, $N > 400$), the false alarm rate using our method is quite comparable with the other two.

Combining the results for the misidentification rate and false alarm rate, we find that our proposed method gains significantly in identification power while giving slightly more false alarms. Overall, the proposed procedure is more robust in identifying the occurring fault and is a more preferable tool for root-cause identification.

Finally, we show the estimation of the fault magnitudes in Table 3. The fault magnitude is estimated under the following setting: $N = 400$ and the noise covariance matrix is fixed at $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon) = 3 \times 10^{-4}$ and $\lambda_{\min}(\mathbf{\Sigma}_\varepsilon) = 10^{-6}$. Each time, we increase the variance level of one element in $\mathbf{f}$ to be ten times larger than the other elements in $\mathbf{f}$ so that this element becomes an outstanding fault. Then, we use Equation (11) to estimate its magnitude. This same simulation is repeated for all p elements in $\mathbf{f}$. Namely, the variations of $\mathbf{f}$ are defined as follows:

$$\operatorname{var}\mathbf{f}(k) = \begin{cases} 2 \times 10^{-2}\,\mathrm{mm}^2 & k = i \\ 2 \times 10^{-3}\,\mathrm{mm}^2 & k \neq i \end{cases}, \quad i = 1 \ldots p.$$

Table 2. False alarm rates of three pattern matching methods

		γ_c (degrees)											
		0		*0.91*		*2.42*		*4.06*		*9.94*		*15.62*	
N	δ	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*	*Mean (%)*	*(Std.) (%)*
50	δ_1	0	(0.000)	0	(0.000)	0	(0.000)	0	(0.000)	5.05	(0.169)	14.83	(0.127)
	δ_2	2.16	(0.183)	2.18	(0.188)	2.13	(0.160)	2.08	(0.158)	2.08	(0.123)	2.02	(0.099)
	δ_3	2.79	(0.182)	6.62	(0.249)	11.44	(0.290)	15.79	(0.099)	16.72	(0.025)	16.85	(0.067)
100	δ_2	0	(0.000)	0	(0.000)	0	(0.000)	0.03	(0.020)	14.75	(0.164)	16.66	(0.015)
	δ_2	11.58	(0.261)	11.66	(0.278)	11.67	(0.311)	11.77	(0.242)	11.58	(0.270)	11.44	(0.292)
	δ_2	12.30	(0.246)	14.90	(0.172)	16.53	(0.036)	16.66	(0.014)	16.67	(0.000)	16.67	(0.000)
400	δ_2	0	(0.000)	0	(0.000)	0.17	(0.052)	10.77	(0.237)	16.67	(0.000)	16.67	(0.000)
	δ_2	16.04	(0.097)	16.02	(0.068)	16.00	(0.064)	16.00	(0.076)	15.89	(0.096)	15.74	(0.130)
	δ_2	16.29	(0.070)	16.66	(0.006)	16.67	(0.000)	16.67	(0.000)	16.67	(0.000)	16.67	(0.000)
1000	δ_2	0	(0.000)	0	(0.000)	7.98	(0.267)	16.56	(0.036)	16.67	(0.000)	16.67	(0.000)
	δ_2	16.25	(0.078)	16.26	(0.088)	16.25	(0.070)	16.22	(0.062)	15.91	(0.092)	15.56	(0.146)
	δ_2	16.39	(0.068)	16.67	(0.000)	16.67	(0.000)	16.67	(0.000)	16.67	(0.000)	16.67	(0.000)

Table 3. The estimated values of the fault magnitudes ($\times 10^{-2}$ mm^2)

	Occurring fault					
	1	*2*	*3*	*4*	*5*	*6*
Magnitude	2.035	2.055	1.995	2.075	2.075	2.015

From Table 3, one can see that the estimated magnitudes for all p faults are very close to the true value, which is 2×10^{-2} mm^2.

3.2.2. *The impact of fault magnitude on* W_k^*

In this subsection, we keep the noise covariance matrix the same as in section 3.2.1, but test three different fault magnitudes in order to illustrate its impact on W_k^*. In all cases, we assume that only the third fault occurs. The three cases are: (i) var $\mathbf{f}(3) = 5 \times 10^{-3}$ mm^2 ; (ii) var $\mathbf{f}(3) = 1 \times 10^{-2}$ mm^2; and (iii) var $\mathbf{f}(3) = 2 \times 10^{-2}$ mm^2. The variances of the other components in $\mathbf{f}$ remain at 2×10^{-3} mm^2.

The W_k^* values were replicated 1000 times for each case. The W_k^* listed in the following tables are the 95th and 99th percentile values of the W_k^* values obtained from all replicates. The sample size N was chosen to be 50, 100, 400, or 1000. When the fault magnitude changes, the worst-case boundary γ_c will change accordingly. The overall results are listed in Table 4, where a zero W_k^* means that the fault symptom vectors always fall inside the corresponding cones for all replicated tests. The values smaller than the corresponding thresholds are marked in bold in the table.

Apparently, when the fault magnitude gets larger, the γ_c due to the perturbation of unstructured noise will decrease. For a given sample size, as the fault magnitude increases, the W_k^* values associated with a non-faulty variable (i.e., $k \neq 3$) increase very fast, whereas the W_3^* values are smaller than the corresponding thresholds and change in a small range.

When the sample size increases, our method is more effective in fault identification, just as expected. Given the same variation of $\mathbf{f}(3)$, when the sample size increases, the value of W_3^* decreases sharply, while other $W_k^*(k \neq 3)$ values will actually increase, which makes the fault identification easier.

When the sample size is relatively small as compared to the dimension of the measurement ($n = 31$) and the fault magnitude is not significantly greater than other elements in $\mathbf{f}$, the fault identification could be more difficult. For instance, when the fault magnitude is 5×10^{-3} mm^2 and $N = 50$ (the first row of the case 1 data for the 99th percentile in Table 4), not only W_3^* but also W_5^* is smaller than $\chi^2_{30}(0.01)$. As a result, we are not able to specifically determine which fault occurred or if both of them did. The phenomenon is more apparent when we use the 95th percentile value of W_k^* (see the first row of the case 1 data for the 95th percentile in Table 4). In order to ensure the effectiveness of the proposed method, the sample size should be large enough. For our case, the proposed method will achieve a satisfactory result when the sample size is larger than 100. When the sample size is 50, the proposed method can identify a fault of a relatively large magnitude, say, when var $\mathbf{f}(3)$ is greater than 1×10^{-2} mm^2.

3.2.3. *The impact of noise magnitude on* W_k^*

This subsection discusses the impact of different noise variances on W_k^*. The variation of $\mathbf{f}$ is specified as,

$$\text{var}\,\mathbf{f}(k) = \begin{cases} 2 \times 10^{-2}\,\text{mm}^2 & k = 3, \\ 2 \times 10^{-3}\,\text{mm}^2 & k \neq 3. \end{cases}$$

Table 4. The W_k^* values for the three studied cases

	95th percentile						*99th percentile*					
N	W_1^*	W_2^*	W_3^*	W_4^*	W_5^*	W_6^*	W_1^*	W_2^*	W_3^*	W_4^*	W_5^*	W_6^*
Case 1 $\gamma_c = 37.41^\circ$												
50	**41.91**	**42.58**	**3.54**	**34.41**	**34.42**	47.00	58.25	66.64	**6.12**	51.37	**49.48**	72.09
100	62.40	56.38	**1.35**	47.68	46.62	66.53	87.33	70.92	**1.52**	68.38	72.43	87.28
400	146.26	141.93	**0**	115.49	114.47	190.83	293.78	291.33	**0**	248.81	240.13	425.06
1000	293.78	291.33	**0**	248.81	240.13	425.06	343.71	323.35	**0**	292.72	277.52	472.26
Case 2 $\gamma_c = 9.94^\circ$												
50	275.34	261.97	**2.90**	202.83	212.02	252.1	357.65	337.71	**5.19**	260.72	298.77	335.63
100	1229.6	1208.8	**1.01**	930.15	895.8	1147.1	528.92	492.62	**3.53**	413.45	416.59	506.08
400	1390.9	1393.9	**1.01**	1066.4	1018	1272.4	1390.9	1393.9	**1.01**	1066.4	1018	1272.4
1000	2846.5	2724.2	**0**	2049	2048.1	2587.1	3186.1	3034	**0**	2317.4	2228.6	2727.6
Case 3 $\gamma_c = 4.06^\circ$												
50	633.65	613.83	**5.26**	522.64	504.61	610.01	810.32	845.16	**8.80**	668.17	639.75	742.43
100	1027	994.36	**2.81**	837.69	808.04	983.02	1275.3	1247.4	**4.87**	959.23	978.96	1152.6
400	3392.6	3221.2	**1.27**	2578.3	2540.1	3065.7	3797.3	3553.8	**3.22**	2947	2834.9	3255.7
1000	7868.3	7641.6	**0.17**	6005.9	5924.2	7063.1	8392.9	8335	**0.17**	6553	6433.2	7426.9

We still assume that the third fault occurs, the same as in the previous sections. Four cases are considered corresponding to different noise covariance structures and the largest and the smallest eigenvalues associated with each case are: (i) $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon) = 9 \times 10^{-5}$ and $\lambda_{\min}(\mathbf{\Sigma}_\varepsilon) = 6 \times 10^{-5}$; (ii) $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon) = 2.4 \times 10^{-4}$ and $\lambda_{\min}(\mathbf{\Sigma}_\varepsilon) = 1.6 \times 10^{-4}$; (iii) $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon) = 3 \times 10^{-4}$ and $\lambda_{\min}(\mathbf{\Sigma}_\varepsilon) = 10^{-6}$; and (iv) $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon) = 4 \times 10^{-4}$ and $\lambda_{\min}(\mathbf{\Sigma}_\varepsilon) = 1 \times 10^{-7}$. The W_k^* values were calculated for $N = 50, 100, 400$, and 1000, respectively. The overall results are listed in Table 5. Again, the values smaller than the corresponding thresholds are marked in bold in the table. We only show the 99th percentile values in this table because the 95th percentile values have the same characteristics when the angle γ_c is not large (in Table 5, γ_c ranges from 0.91° to 5.42°).

Comparing cases (i) and (ii) where the $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)/\lambda_{\min}(\mathbf{\Sigma}_\varepsilon)$ ratios are the same, we find that the perturbation boundary γ_c will increase from 0.91° in case (i) to 2.42° in case (ii) when the ratio of $\lambda_{\mathbf{f},i}/\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)$ decreases from 188.3 in case (i) to 70.6 in case (ii). This effect is similar to the situation that a fault magnitude decreases while the noise magnitude is kept the same. When γ_c increases, meaning the optimization range will be accordingly enlarged, W_k^* will decrease. Moreover, when the sample size increases, W_3^* will decrease but the other $W_k^*(i \neq 3)$ values will increase. This result indicates that a larger ratio of $\lambda_{\mathbf{f},i}/\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)$ or a larger sample size will make the fault identification easier.

Comparing cases (iii) and (iv), where the ratio of $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)/\lambda_{\min}(\mathbf{\Sigma}_\varepsilon)$ are made significantly different (300 in case (iii) compared to 4000 in case (iv)), we find that an increasing $\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)/\lambda_{\min}(\mathbf{\Sigma}_\varepsilon)$ will enlarge the perturbation boundary γ_c (4.06° in case (i) compared to 5.42° in case (ii)). Accordingly, W_3^* will decrease for a given sample size and the probability to identify a fault using our method becomes higher. As always, a larger sample size will lead us to a better result.

Table 5. The W_k^* values for the four studied cases

N	W_1^*	W_2^*	W_3^*	W_4^*	W_5^*	W_6^*
Case 1 $\gamma_c = 0.91°$						
50	887.92	832.09	**13.619**	702.69	687.22	813.77
100	1215.2	1322	**12.012**	1033.3	982.78	1127.5
400	3859.8	3803.6	**10.777**	3045.6	3176.9	3449.1
1000	8612.6	8625.4	**8.5097**	7219.2	7491.1	7814.9
Case 2 $\gamma_c = 2.42°$						
50	821.4	739.83	**18.294**	607.47	662.58	697.24
100	1192.7	1113.1	**10.858**	939.06	982.8	1099.8
400	3406.8	3471.7	**5.8144**	2818.8	2803.8	3236.4
1000	7907.2	7752.8	**3.5624**	6299.9	6282.9	7241.9
Case 3 $\gamma_c = 4.06°$						
50	810.32	845.16	**8.8074**	668.17	639.75	742.43
100	1275.3	1247.4	**4.8704**	959.23	978.96	1152.6
400	3797.3	3553.8	**3.2185**	2947	2834.9	3255.7
1000	8392.9	8335	**0.1719**	6553	6433.2	7426.9
Case 4 $\gamma_c = 5.42°$						
50	834.96	737.71	**6.1689**	684.41	673.56	673.56
100	1203.8	1204.9	**3.6969**	979.99	986.79	1093
400	3598.9	3405.1	**0.0477**	2880.9	2921.6	3199.9
1000	8040.7	7823.3	**0**	6569.7	6508.8	7173.8

4. Conclusions

The pattern matching technique is one of the two most widely used approaches for root-cause identification. It has a several apparent advantages including: (i) there is a clear geometric explanation of fault signatures or fault patterns; (ii) the results can thus be easily visualized and physicaly interpretated; and (iii) it is intuitive and easy for practitioners to implement and execute. We have developed a procedure to make the pattern matching method more robust under the perturbations due to unstructured noise and sampling uncertainties. Case studies have been conducted to illustrate the effectiveness of this method. It is demonstrated that the identification capability using the proposed procedure is significantly improved. Also, in the case study, the impact of different noise covariance structures, fault magnitudes, and sample sizes is discussed. This method can be used for a variation source identification in manufacturing processes and will reduce product variability. One point we need to make clear here is that the technique presented in this paper is a follow-up step of fault detection, i.e., after a fault is detected, we can decide if 1-out-of-p known faults occurs or not by using the presented technique. Fault detection itself is an interesting problem (Apley and Shi, 2001) and is not covered in this paper.

There are still some open issues worthy of future investigation. Firstly, this method is developed by assuming that one fault occurs at a time. Extension of the methodology to multiple faults cases is under investigation. A second issue concerns the fault isolation capability of the method. When the angle between two fault signature vectors (i.e., two column vectors of **A**) is small, the method may not be able to distinguish them and isolate the true fault source. The isolation condition under which all faults are uniquely identifiable is worth further study. In addition, how to identify the new fault(s) in the processes, referred to as the *novelty identification problem*, is a very interesting problem. Through investigating this problem, we can extend our knowledge on process faults and expand the fault domain that we can identify.

Acknowledgements

The authors gratefully acknowledge the financial support of NSF under grants DMI-0217481, DMI-0322147, and DMI-0348150, and from the State of Texas Advanced Technology Program under grant 000512-0237-2003. The authors would like to thank the Editors and reviewers for

their insightful comments and suggestions, which have significantly improved the paper's quality and readability.

References

Apley, D.W. and Shi, J. (1998) Diagnosis of multiple fixture faults in panel assembly. *ASME Journal of Manufacturing Science and Engineering*, **120**, 793–801.

Apley, D.W. and Shi, J. (2001) A factor-analysis method for diagnosing variability in multivariate manufacturing processes.*Technometrics*, **43**, 84–95.

Ceglarek, D. and Shi, J. (1996) Fixture failure diagnosis for autobody assembly using pattern recognition. *ASME Journal of Engineering for Industry*, **188**, 55–65.

Chang, M. and Gossard, D.C. (1998) Computational method for diagnosis of variation-related assembly problem. *International Journal of Production Research*, **36**, 2985–2995.

Ding, Y., Ceglarek, D. and Shi, J. (2002a) Fault diagnosis of multistage manufacturing processes by using state space approach. *Transactions of the ASME. Journal of Manufacturing Science and Engineering*, **124**, 313–322.

Ding, Y., Shi, J. and Ceglarek, D. (2002b) Diagnosability analysis of multi-station manufacturing processes. *Transactions of the ASME. Journal of Dynamic Systems, Measurement, and Control*, **124**, 1–13.

Ding, Y., Zhou, S. and Chen, Y. (2005) A comparison of process variation estimators for in-process dimensional measurements and control. *Transactions of the ASME. Journal of Dynamic Systems, Measurement and Control,* **127**, 69–79.

Djurdjanovic, D. and Ni, J. (2001) Linear state space modeling of dimensional machining errors. *Transactions of NAMRI/SME*, **XXIX**, 541–548.

Golub, G.H. and Van Loan, C.F. (1996) *Matrix Computations*, 3rd edn., The Johns Hopkins University Press, Baltimore, MD.

Jin, J. and Shi, J. (1999) State space modeling of sheet metal assembly for dimensional control. *ASME Journal of Manufacturing Science and Engineering*, **121**, 756–762.

Johnson, R.A. and Wichern, D.W. (1998) *Applied Multivariate Statistical Analysis*, 4th edn., Prentice-Hall, NJ.

McCulloch, C. and Searle, S.R. (2001) *Generalized, Linear, and Mixed Models*, Wiley, New York, NY.

Montgomery, D.C. (2005) *Introduction to Statistical Quality Control*, 5th edn., Wiley, New York, NY.

Murihead, R. (1982) *Aspects of Multivariate Statistical Theory*, Wiley, New York, NY.

Rong, Q., Ceglarek, D. and Shi, J. (2000) Dimensional fault diagnosis for compliant beam structure assemblies. *Transactions of the ASME. Journal of Manufacturing Science and Engineering*, **122**, 773–780.

Searle, S.R., Casella, G. and McCulloch, C.E. (1992) *Variance Components*, Wiley, New York, NY.

Taha, H.A. (1992) *Operations Research, An Introduction*, 5th edn., Prentice-Hall, NJ. Jersey.

Zhou, S., Chen, Y. and Shi, J. (2004) Root cause estimation and statistical testing for quality improvement of multistage manufacturing processes. *IEEE Transactions on Automation Science and Engineering*, **1**, 73–83.

Zhou, S., Ding, Y., Chen, Y. and Shi, J. (2003a) Diagnosability study of multistage manufacturing process based on linear mixed-effects models. *Technometrics*, **45**, 312–325.

Zhou, S., Huang, Q. and Shi, J. (2003b) State space modeling for dimensional monitoring of multistage machining process using differential motion vector. *IEEE Transactions on Robotics and Automation*, **19**, 296–308.

Appendix

In order to prove that $\sin(\Delta\theta) = \text{dist}(\text{span}\{\mathbf{a}_i\}, \text{span}\{\mathbf{v}_i^{(1)}\})$ we need the concepts of orthogonal projection and the distance between two subspaces.

Orthogonal projection: Let $\mathbf{S} \subseteq R^n$ be a subspace, and $\mathbf{P} \in R^{n\times n}$ is the orthogonal projection onto $\mathbf{S}$ if range $(\mathbf{P}) = \mathbf{S}$, $\mathbf{P}^2 = \mathbf{P}$, and $\mathbf{P}^T = \mathbf{P}$. It is known that if $\mathbf{v} \in R^n$, and $\mathbf{v}^T\mathbf{v} = 1$, then $\mathbf{P} = \mathbf{v}\mathbf{v}^T$ is the orthogonal projection onto $\mathbf{S} = \text{span}\{\mathbf{v}\}$, where $\text{span}\{\cdot\}$ represents the subspace spanned by $\mathbf{v}$.

Distance between subspaces: Suppose $\mathbf{S}_1$ and $\mathbf{S}_2$ are subspaces of R^n, and they have the same dimension, then the distance between these two subspaces is $\text{dist}(\mathbf{S}_1, \mathbf{S}_2) = \|\mathbf{B}_1 - \mathbf{B}_2\|_2$, where $\mathbf{B}_1$, $\mathbf{B}_2$ are the orthogonal projection to $\mathbf{S}_1$, $\mathbf{S}_2$ respectively.

Proof. Given $\|\mathbf{a}_i\|_2 = \|\mathbf{v}_i^{(1)}\|_2 = 1$, it is obvious that $\sin(\Delta\theta) = \sqrt{1 - \cos^2(\Delta\theta)} = \sqrt{1 - (\mathbf{a}_i^T\mathbf{v}_i^{(1)})^2}$.

We know $\mathbf{a}_i\mathbf{a}_i^T$ and $\mathbf{v}_i^{(1)}\mathbf{v}_i^{(1)T}$ are the orthogonal projections on to the subspaces $\text{span}\{\mathbf{a}_i\}$ and $\text{span}\{\mathbf{v}_i^{(1)}\}$ respectively, hence, $\text{dist}(\text{span}\{\mathbf{a}_i\}, \text{span}\{\mathbf{v}_i^{(1)}\}) = \|\mathbf{a}_i\mathbf{a}_i^T - \mathbf{v}_i^{(1)}\mathbf{v}_i^{(1)T}\|_2$. Suppose that $\mathbf{X}$ and $\mathbf{Z}$ are two $p \times p$ orthogonal matrices and $\mathbf{a}_i \in \mathbf{X}, \mathbf{v}_i^{(1)} \in \mathbf{Z}$ so that $\mathbf{X} = [\mathbf{a}_i, \mathbf{X}_2]$ and $\mathbf{Z} = [\mathbf{v}_i^{(1)}, \mathbf{Z}_2]$. Then we have that:

$$\begin{aligned}\text{dist}\big(\text{span}\{\mathbf{a}_i\}, \text{span}\{\mathbf{v}_i^{(1)}\}\big) &= \big\|\mathbf{a}_i\mathbf{a}_i^T - \mathbf{v}_i^{(1)}\mathbf{v}_i^{(1)T}\big\|_2 \\ &= \big\|\mathbf{X}\big(\mathbf{a}_i\mathbf{a}_i^T - \mathbf{v}_i^{(1)}\mathbf{v}_i^{(1)T}\big)\mathbf{Z}\big\|_2 = \left\|\begin{bmatrix} 0 & \mathbf{a}_i^T\mathbf{Z}_2 \\ -\mathbf{X}_2^T\mathbf{v}_i^{(1)} & \mathbf{0}\end{bmatrix}\right\|_2.\end{aligned}$$

Note that vectors $\mathbf{a}_i^T\mathbf{Z}_2$ and $\mathbf{X}_2^T\mathbf{v}_i^{(1)}$ are parts of the orthogonal matrix:

$$\mathbf{T} = \begin{bmatrix}\mathbf{a}_i^T\mathbf{v}_i^{(1)} & \mathbf{a}_i^T\mathbf{Z}_2 \\ \mathbf{X}_2^T\mathbf{v}_i^{(1)} & \mathbf{X}_2^T\mathbf{Z}_2\end{bmatrix} = \begin{bmatrix}\mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22}\end{bmatrix}.$$

Thus, we get.

$$\begin{aligned}\|\mathbf{T}_{21}\|_2^2 &= \max_{\|\mathbf{x}\|_2=1}\|\mathbf{T}_{21}\mathbf{x}\|_2^2 = 1 - \min_{\|\mathbf{x}\|_2=1}\|\mathbf{T}_{11}\mathbf{x}\|_2^2 \\ &= 1 - \sigma_{\min}(\mathbf{T}_{11})^2 = 1 - \big(\mathbf{a}_i^T\mathbf{v}_i^{(1)}\big)^2.\end{aligned}$$

Likewise, using $\mathbf{T}^T$, it is easy to show that $\|\mathbf{T}_{12}\|_2^2 = 1 - \sigma_{\min}(\mathbf{T}_{11})^2 = 1 - (\mathbf{a}_i^T\mathbf{v}_i^{(1)})^2$, hence $\|\mathbf{T}_{21}\|_2^2 = \|\mathbf{T}_{12}\|_2^2 = 1 - (\mathbf{a}_i^T\mathbf{v}_i^{(1)})^2$. Here $\sigma(\cdot)$ means the set of the singular values.

Since $\text{dist}(\text{span}\{\mathbf{a}_i\}, \text{span}\{\mathbf{v}_i^{(1)}\}) = \|\mathbf{T}_{21}\|_2^2 = \|\mathbf{T}_{12}\|_2^2 = 1 - (\mathbf{a}_i^T\mathbf{v}_i^{(1)})^2$ (refer to Theorem 2.6.1 of Golub and Van Loan (1996, p. 76), $\|\mathbf{T}_{21}\|_2^2 = \|\mathbf{T}_{12}\|_2^2 = 1 - (\mathbf{a}_i^T\mathbf{v}_i^{(1)})^2$ is $\sin^2(\Delta\theta)$. ■

Biographies

Zhiguo Li is a Ph.D. student in the Department of Industrial and Systems Engineering at University of Wisconsin-Madison. His research interests focus on variation source identification in large complex systems, statistical modeling and reliability analysis of event sequence data.

Shiyu Zhou is an assistant professor in the Department of Industrial and Systems Engineering at the University of Wisconsin-Madison. He got his B.S. and M.S. in Mechanical Engineering at University of Science and Technology of China in 1993 and 1996 respectively, and got his Master in Industrial Engineering and Ph.D. in Mechanical Engineering at the University of Michigan in 2000. Dr. Zhou's research interests are the in-process quality and productivity improvement methodologies by integrating statistics, system and control theory, and engineering knowledge. The objective is to achieve automatic process monitoring, diagnosis, compensation, and their implementation in various manufacturing processes. His research is sponsored by National Science Foundation, Department of Energy, NIST-ATP, and industries. He is a recipient of the CAREER Award from the National Science Foundation in 2006. Dr. Zhou is a member of IIE, INFORMS, ASME, and SME.

Yu Ding received B.S. degrees in Precision Engineering from University of Science and Technology of China in 1993, M.S. in Precision Instruments from Tsinghua University, China in 1996, M.S. in Mechanical Engineering from the Pennsylvania State University in 1998, and Ph.D. in Mechanical Engineering from the University of Michigan in 2001. He is currently an assistant professor in the Department of Industrial and Systems Engineering at Texas A&M University. Dr. Ding's research interests are in the area of quality engineering and applied statistics, including in-process variation diagnosis, diagnosability analysis of distributed sensor systems, optimal sensor system design, and process-oriented robust design and tolerancing. His current research is sponsored by the National Science Foundation, Nokia, and the State of Texas Higher Education Coordinating Board. Dr. Ding received a number of awards for his work, including the CAREER Award from the National Science Foundation in 2004 and the Best Paper Award from the ASME Manufacturing Engineering Division in 2000. Dr. Ding currently serves as a department editor of *IIE Transactions* and an associate editor of *IEEE Transactions on Automation Science and Engineering*. Dr. Ding is a member of IIE, IEEE, INFORMS, and ASME.