IIE Transactions (2007) **39**, 691–701

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170600899573

Identifying and visualizing nonlinear variation patterns in multivariate manufacturing data

DANIEL W. APLEY[1,*] and FENG ZHANG[2]

[1] *Department of Industrial Engineering and Management Science, Northwestern University, Evanston, IL 60208-3119, USA*
Email: apley@northwestern.edu
[2] *Fairchild Semiconductor, 82 Running Hill Road, South Portland, ME 04106, USA*
Email: Feng.Zhang@fairchildsemi.com

Received March 2004 and accepted October 2005

In modern manufacturing processes, it is not uncommon to have hundreds, or even thousands, of different process variables that are measured and recorded in databases. The large quantities of multivariate measurement data usually contain valuable information on the major sources of variation that contribute to the final product or process variability. The focus of this work is on variation sources that result in complex, nonlinear variation patterns in the data. We propose a model for representing nonlinear variation patterns and a method for blindly identifying the patterns, based on a sample of measurement data, with no prior knowledge of the nature of the patterns. The identification method is based on principal curve estimation, in conjunction with a data preprocessing step that makes it suitable for high dimensional data. We also discuss an approach for interactively visualizing the nature of the identified variation patterns, to aid in identifying and eliminating the major root causes of manufacturing variation.

Keywords: Principal curves, principal component analysis, blind identification, variation patterns, statistical process control

1. Introduction

Recent advances in in-process measurement and data collection technologies in manufacturing settings, are allowing hundreds and even possibly thousands of different process variables to be measured, often for 100% of the parts being produced. Companies invest in measurement technology and generate large quality-related databases with the expectation that this will somehow aid in efforts to systematically identify and eliminate the major root causes of manufacturing variation. Effectively transforming the large sets of measurement data into extracted knowledge that is useful for variation reduction efforts, however, is a challenging data mining problem.

We investigate this problem following the same paradigm considered in Apley and Lee (2003), in which one views each major variation source as causing a distinct variation pattern in the data, spreading across any or all of the different variables that are measured. Each variation pattern represents the interrelated manner in which the different variables vary from part to part, due to the effects of a particular variation source. The data mining objective becomes one of unsupervised learning, in which we seek to discover the nature of any variation pattern that happens to be present in the data. We refer to this as *blindly* identifying the variation patterns, in the sense that we are not attempting to recognize the presence of premodeled or pretrained patterns. Rather, we seek to identify the nature of the patterns based only on a sample of data, with no prior training or modeling required. After blindly identifying the nature of a variation pattern, the results can be graphically illustrated in order to facilitate root cause identification, which we illustrate with examples in later sections.

A number of techniques have been recently developed for blindly identifying the nature of variation patterns in large sets of manufacturing measurement data (e.g., Apley and Shi (2001), Apley and Lee (2003), and Lee and Apley (2004)). All of these approaches assume that the variation patterns can be represented using linear models, however. Although linear models are reasonably versatile, there are many situations in which nonlinear models are required to represent variation patterns, such as in the following example from autobody assembly. Figure 1 illustrates schematically the rear liftgate opening of a sports utility vehicle and shows the locations at which six cross-car (left/right) dimensional measurements on the left and right bodysides are taken (denoted x_1 through x_6). The measurements are obtained automatically via in-process laser measurement, so that every autobody is measured (for a

*Corresponding author

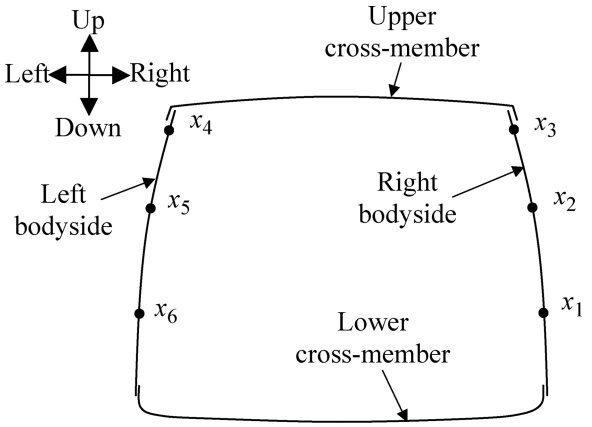

Fig. 1. Measurement layout on the liftgate opening of an autobody. The fore direction is pointing into the page.

more detailed description of the assembly process and measurement technology, refer to Ceglarek and Shi (1996) or Apley and Shi (1998; 2001)). Although almost 200 different dimensional features were measured for each autobody, for simplicity we illustrate with only the six measurements shown in Fig. 1. The nonlinear variation pattern described in the following paragraphs primarily affected only the liftgate region of the autobody.

The assembly process is relatively complex and involves hundreds of different assembly stations and thousands of different tooling elements. When a tooling element breaks, wears, malfunctions, or is simply not designed properly, this often results in a distinct variation pattern in the dimensional measurement data. Figure 2 illustrates this with scatter plots of pairs of the six variables over a sample of 100 measured autobodies. The measurements are deviations from nominal, in units of millimeters. A positive measurement represents deviation to the right. Although the relationship between x_2 and x_3 appears linear, the scatter plots for x_2/x_5 and x_3/x_4 clearly illustrate that the variation pattern is nonlinear. Moreover, this nonlinear pattern appears to be approximately piecewise linear with only two segments (i.e., pieces).

One potential root cause for the variation pattern is due to a fixturing problem when locating the right bodyside in the framing station (a major assembly station in which the left and right bodysides are joined to the underbody and a set of upper cross-members). When the right bodyside deviates by only a small amount to the right, it has no affect on the left bodyside. When the right bodyside deviates by a larger amount to the right, however, it begins to interfere with the upper cross-member. The upper cross-member then interferes with the left bodyside by also pulling the left bodyside towards the right.

In situations like that depicted in Fig. 2, linear models are inadequate to represent the nonlinear relationship between the different variables. One potential method for treating nonlinear variation patterns is based on the notion

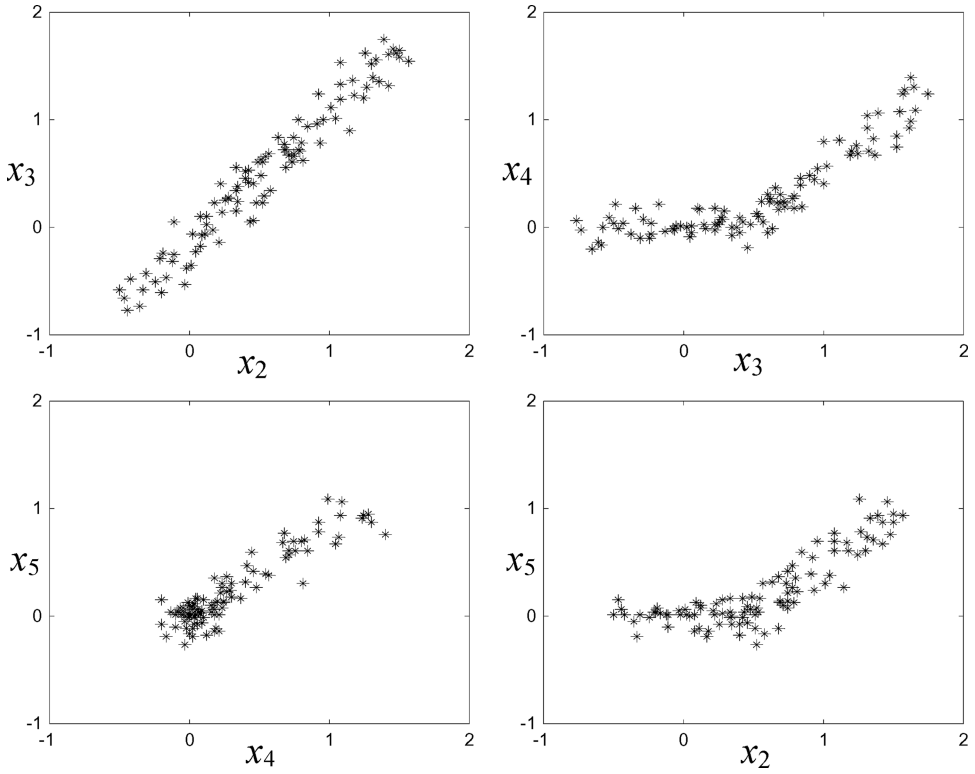

Fig. 2. Scatter plots of data from 100 measured autobodies exhibiting a nonlinear variation pattern.

of a principal curve (Hastie and Stuetzle, 1989), which is a nonlinear generalization of Principal Components Analysis (PCA). Broadly speaking, a principal curve is defined as a one-dimensional curve that passes through the middle of the distribution of higher dimensional data. Principal curve estimation is relatively robust and has been applied to a variety of nonlinear data analysis problems (Tibshirani, 1996; Chang and Ghosh, 2001; Delicado, 2001). However, most applications of principal curves have been for relatively low dimensional data, especially for two-dimensional image processing (e.g., Banfield and Raftery (1992), Kégl *et al.*, (2000), and also Chang and Ghosh (2001)). For the high dimensional data often encountered in manufacturing, principal curve estimation becomes inefficient. Because of this, we use a common data preprocessing step that involves linear PCA to first reduce the dimensionality of the problem. This not only reduces the computational complexity, but also improves the estimation accuracy by filtering out a substantial portion of the noise (i.e., random variations that are not due to any systematic pattern). One advantage of the principal curve approach is that it lends itself well to visualizing the blindly identified variation patterns. Effective visualization of a variation pattern is crucial for identifying the root cause of the variation. In a later section, we illustrate the visualization approach with an example in which the data are extremely high dimensional, representing point cloud data from laser-scanned stamped panels.

The remainder of the paper is organized as follows. Section 2 discusses the model we use to represent nonlinear manufacturing variation patterns. Section 3 provides some background on linear PCA and nonlinear principal curves. Although PCA is a common first step for reducing dimensionality in data mining applications, including principal curve estimation (e.g., Delicado and Huerta (2003)), the amount of information that is lost ranges from nothing at all to perhaps a non-negligible amount. In Section 4, we argue that for our application virtually no information is lost. In Section 5, we describe the algorithm for principal curve estimation with the PCA preprocessing step. Examples illustrating the approach are provided in Sections 5 and 6. Section 7 includes a Monte Carlo analysis demonstrating the performance improvement when one uses linear PCA to first reduce the dimensionality of the problem

2. Representing linear and nonlinear variation patterns

In most of the previous work on identifying variation patterns in manufacturing data (e.g., Apley and Shi (2001) and also Apley and Lee (2003)), the following linear model was used to represent the pattern. Let $\mathbf{X} = [X_1, X_2, \ldots, X_d]^T$ be a $d \times 1$ random vector that represents a set of d measured characteristics from the product or process. Let $\{\mathbf{x}_i: i = 1, 2, \ldots, N\}$ be a sample of N observations of $\mathbf{X}$. In autobody assembly, for example, $\mathbf{X}$ would represent the vector of all measured dimensional characteristics across a given autobody, and N would denote the number of autobodies in the sample. If there is only a single variation source present in the manufacturing process over a particular sample, the linear model takes the form:

$$\mathbf{X} = \mathbf{c}t + \mathbf{W}, \tag{1}$$

where $\mathbf{c}$ is a $d \times 1$ constant vector and t is a scalar random variable that is scaled (without loss of generality) to have unit variance. The noise $\mathbf{W} = [W_1, W_2, \ldots, W_d]^T$ is a $d \times 1$ zero-mean random vector that is independent of t.

The interpretation of the model is that the effects of the variation source on the measurements are represented by the elements of $\mathbf{c}$. The vector $\mathbf{c}$ therefore indicates the spatial nature of the variation pattern, in terms of the resulting interrelationships between the different measured variables. The random vector $\mathbf{W}$ represents the aggregated effects of measurement noise and any unsystematic variation that is not attributed to the major variation source. Unless otherwise noted, it is assumed throughout that the covariance matrix of $\mathbf{W}$ is $\Sigma_{\mathbf{w}} = \sigma_{\mathbf{w}}^2 \mathbf{I}$, a scalar multiple of the identity matrix. Apley and Lee (2003) discussed approaches for dealing with more general noise covariance structure, which also apply to the present case of nonlinear patterns.

Although linear models of this form are relatively common for representing manufacturing variation (Barton and Gonzalez-Barreto, 1996; Ceglarek and Shi, 1996; Jin and Shi, 1999; Ding *et al.*, 2002; Zhou *et al.*, 2003), they cannot describe more complex nonlinear phenomena. In analogy with the linear model (1), we propose the following model for representing nonlinear variation patterns:

$$\mathbf{X} = \mathbf{f}(t) + \mathbf{W}, \tag{2}$$

where t and $\mathbf{W}$ are as in the linear model, and $\mathbf{f}(\cdot)$ is some general vector-valued nonlinear function that represents the spatial nature of the nonlinear variation pattern. For the variation pattern shown in Fig. 2, $\mathbf{f}(\cdot)$ would be (at least approximately) a piecewise linear function with two pieces.

In the linear situation, the objective is to blindly estimate $\mathbf{c}$, given a sample of data. Apley and Shi (2001) and Apley and Lee (2003) provide examples of this and discuss how graphical illustrations of the estimated $\mathbf{c}$ can aid in identifying the root cause of the variation. In the nonlinear case, the objective is to blindly estimate the entire function $\mathbf{f}(\cdot)$, a suitable illustration of which can aid in root cause identification. The remainder of this paper focuses on how to estimate $\mathbf{f}(\cdot)$.

3. PCA and principal curves

Let $\boldsymbol{\mu}_{\mathbf{x}}$ and $\boldsymbol{\Sigma}_{\mathbf{x}}$ denote the mean vector and covariance matrix, respectively, of $\mathbf{X}$. For linear variation patterns, PCA can be used directly to provide an estimate of $\mathbf{c}$. PCA involves the analysis of the eigenvectors and eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{x}}$ (Johnson and Wichern, 2002). Denote the eigenvalues and

corresponding eigenvectors by $\{\lambda_j : j = 1, 2, \ldots, d\}$ and $\{\mathbf{z}_j : j = 1, 2, \ldots, d\}$, respectively, where the eigenvalues are arranged in descending order. It is well known that when a single linear variation pattern is present, the dominant eigenvector $\mathbf{z}_1$ is a scaled version of $\mathbf{c}$ (Ceglarek and Shi, 1996; Apley and Shi 2001). Thus, the dominant eigenvector of the sample covariance matrix can be used as an estimate of $\mathbf{c}$.

One related fundamental property of PCA is that the straight line $\boldsymbol{\mu}_\mathbf{x} + t\mathbf{z}_1$ (with t viewed as a scalar parameter that determines the distance along the line) provides the best one-dimensional linear approximation to the distribution of $\mathbf{X}$, in the sense of minimizing the mean square (orthogonal) distance between $\mathbf{X}$ and the approximating line. Principal curves are a natural nonlinear generalization of this concept. Because $\mathbf{f}(t)$ in Equation (2) satisfies the definition of a principal curve stated below, we also use it to denote a principal curve. According to the definition of Hastie and Stuetzle (1989) (hereafter referred to as HS) a one-dimensional curve $\mathbf{f}(t) = [f_1(t), f_2(t), \ldots, f_d(t)]^T$ in d-dimensional space is a principal curve of (the distribution of) $\mathbf{X}$ if:

1. $\mathbf{f}(t)$ does not intersect itself;
2. $\mathbf{f}(t)$ has finite length inside any bounded subset of $\Re^d$; and
3. $\mathbf{f}(t) = E(\mathbf{X} \mid t_\mathbf{f}(\mathbf{X}) = t)$, where $E(\cdot|\cdot)$ denotes a conditional expectation, and the projection index $t_\mathbf{f}(\mathbf{x})$ is defined as

$$t_\mathbf{f}(\mathbf{x}) = \sup\{t : ||\mathbf{x} - \mathbf{f}(t)|| = \inf_u \|\mathbf{x} - \mathbf{f}(u)\|\}. \qquad (3)$$

HS showed that no infinitesimally small smooth perturbation to a principal curve will decrease $E||\mathbf{X} - \mathbf{f}(t_\mathbf{f}(\mathbf{X}))||^2$. In this sense principle curves generalize the minimum mean square distance property of the linear PCA approximation to the distribution of $\mathbf{X}$. In the hypothetical case that the distribution of $\mathbf{X}$ is known, the HS algorithm for constructing principal curves iterates over the following steps:

Step 0. Initialize $\mathbf{f}^{(0)}(t) = \boldsymbol{\mu}_\mathbf{x} + t\mathbf{z}_1$, and set $j = 1$.
Step 1. (Expectation.) Define $\mathbf{f}^{(j)}(t) = E[\mathbf{X}|t_{\mathbf{f}^{(j-1)}}(\mathbf{X}) = t]$ for all t.
Step 2. (Projection.) For each $\mathbf{x} \in \Re^d$, set $t_{\mathbf{f}^{(j)}}(\mathbf{x}) = \max\{t : \|\mathbf{x} - \mathbf{f}^{(j)}(t)\| = \min_u \|\mathbf{x} - \mathbf{f}^{(j)}(u)\|\}$.
Step 3. Compute $\Delta(\mathbf{f}^{(j)}) = E\|\mathbf{X} - \mathbf{f}^{(j)}(t_{\mathbf{f}^{(j)}}(\mathbf{X}))\|^2$. If $|\Delta(\mathbf{f}^{(j)}) - \Delta(\mathbf{f}^{(j-1)})| < \textit{threshold}$, then stop. Otherwise, let $j = j + 1$ and go to Step 1.

In practice, the theoretical distribution of $\mathbf{X}$ is unknown and the HS algorithm must be implemented on a sample $\{\mathbf{x}_i : i = 1, 2, \ldots, N\}$ of data. The actual HS algorithm starts with $\hat{\mathbf{f}}^{(0)}(t) = \hat{\boldsymbol{\mu}}_\mathbf{x} + t\hat{\mathbf{z}}_1$ and the expectations in Steps 1 and 3 are replaced by some form of locally weighted sample averages. Throughout, the overscore symbol "^" will be used to denote an estimate of a quantity.

4. Principal curves in lower dimensional linear varieties

If we are to effectively use principal curve concepts in the identification of nonlinear variation patterns in the high dimensional data typically encountered in manufacturing, we must first reduce the dimensionality of the problem using linear PCA as a preprocessing step, as we will describe in the following section. In the present section, we show that no information is lost in the dimensionality reduction step, if $\mathbf{f}(t)$ lies in a linear variety (a translated linear subspace) of dimension $r < d$. We also argue that there are many practical manufacturing situations in which nonlinear variation patterns lie in lower dimensional subspaces, in particular when $\mathbf{f}(t)$ can be closely approximated as a piecewise linear curve.

Let $\mathbf{M}$ denote some r-dimensional subspace of $\Re^d$, and let $\mathbf{a}_0 + \mathbf{M}$ denote the r-dimensional linear variety that results from translating $\mathbf{M}$ by some constant vector $\mathbf{a}_0$. Suppose that $\mathbf{f}(t)$ lies in $\mathbf{a}_0 + \mathbf{M}$ for all t and that the noise covariance is $\Sigma_\mathbf{w} = \sigma_\mathbf{w}^2 \mathbf{I}$. Suppose also that no other linear variety in which $\mathbf{f}(t)$ lies has a dimension smaller than r. We show in the Appendix that in this case,

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > \sigma_\mathbf{w}^2 = \lambda_{r+1} = \lambda_{r+2} \ldots = \lambda_d, \qquad (4)$$

and

$$\mathbf{a}_0 + \mathbf{M} = \boldsymbol{\mu}_\mathbf{x} + \text{span}\ \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_r\}. \qquad (5)$$

The significance is that when $\mathbf{f}(t)$ lies in an r-dimensional linear variety, PCA can be used to identify the linear variety. The linear variety is given by the span of the first r eigenvectors, translated by $\boldsymbol{\mu}_\mathbf{x}$, where r is equal to the number of dominant eigenvalues. Although Equation (4) applies to the theoretical covariance matrix $\Sigma_\mathbf{x}$, in practice PCA is conducted on a sample covariance matrix. In this case, it may not be clear what value should be chosen for r. There are, however, a number of statistical methods that have been proposed for estimating the number of dominant eigenvalues in this situation. The most common methods use either likelihood-based or information-based criteria, and involve the ratio of the arithmetic and geometric averages of the small eigenvalues. The reader is referred to Apley and Shi (2001), and the references therein, for a detailed discussion on the various methods for estimating the number of dominant eigenvalues, including their relative performances in typical manufacturing scenarios.

Although the algorithm described in Section 5 fits a principal curve within the r-dimensional space of dominant PCA scores, it is conceptually helpful to think of fitting a principal curve within the original d-dimensional space, but using a "PCA-filtered" version of the data to reduce the effects of the noise: Consider the $d \times r$ matrix $\mathbf{Z} = [\mathbf{z}_1 \mathbf{z}_2 \ldots \mathbf{z}_r]$ and the $d \times d$ matrix $\mathbf{P}_\mathbf{z} = \mathbf{Z}\mathbf{Z}^T$, which is the projection operator onto $\mathbf{M}$. Define the filtered version $\mathbf{X}_\mathbf{z} = \mathbf{P}_\mathbf{z}(\mathbf{X} - \boldsymbol{\mu}_\mathbf{x}) + \boldsymbol{\mu}_\mathbf{x}$ of the random vector $\mathbf{X}$.

Then

$$\mathbf{X_z} = \mathbf{P_z}(\mathbf{X} - \boldsymbol{\mu}_\mathbf{x}) + \boldsymbol{\mu}_\mathbf{x} = \mathbf{P_z}(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x} + \mathbf{W}) + \boldsymbol{\mu}_\mathbf{x}$$
$$= \mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x} + \mathbf{P_z}\mathbf{W} + \boldsymbol{\mu}_\mathbf{x} = \mathbf{f}(t) + \mathbf{W_z}, \quad (6)$$

where the third equality follows from the fact that $\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x} \in \mathbf{M}$, and $\mathbf{W_z} = \mathbf{P_z}\mathbf{W}$ is the filtered version of the noise, obtained by projecting $\mathbf{W}$ onto the r-dimensional subspace $\mathbf{M}$.

Equation (6) implies that by filtering the data using the results of PCA, the principal curve component $\mathbf{f}(t)$ of the model (2) is unchanged, whereas the noise component $\mathbf{W}$ is reduced by an amount that depends on d and r. To quantify the extent to which the noise is reduced, consider the total variance $E[\mathbf{W}^T\mathbf{W}]$ as a measure of the noise level. Before filtering, the total noise variance is $E[\mathbf{W}^T\mathbf{W}] = \text{trace}\ \{E[\mathbf{W}\mathbf{W}^T]\} = \text{trace}\ \{\sigma_\mathbf{w}^2\ \mathbf{I}\} = d\sigma_\mathbf{w}^2$. After filtering, the total variance of the filtered noise is $E[\mathbf{W}_\mathbf{z}^T\mathbf{W}_\mathbf{z}] = \text{trace}\ \{E[\mathbf{W_z}\ \mathbf{W}_\mathbf{z}^T]\} = \text{trace}\{\mathbf{P_z}\Sigma_\mathbf{w}\ \mathbf{P_z}\} = \boldsymbol{\sigma}_\mathbf{w}^2\ \text{trace}\ \{\mathbf{P_z}\mathbf{P_z}\} = r\sigma_\mathbf{w}^2$. Thus, the ratio of total noise variance before and after filtering is d/r. When $\mathbf{f}(t)$ lies in a linear variety of dimension much smaller than d, the reduction in noise variance will be substantial. The simulation results presented in Section 7 demonstrate that this improves the principal curves estimation accuracy.

In what type of situations might we expect $\mathbf{f}(t)$ to lie in a lower dimensional linear variety? The autobody example introduced in Section 1 is one situation, because the principal curve $\mathbf{f}(t)$ appears to be piecewise linear with two linear segments. More generally, suppose $\mathbf{f}(t)$ is piecewise linear with p linear segments, which is illustrated in Fig. 3 for the case that $p = 4$ and $d = 3$. Although we illustrate this with $p > d$, the primary utility of the approach will be for situations in which $p \ll d$. The following arguments show that a piecewise linear $\mathbf{f}(t)$ lies in a linear variety whose dimension is equal to the number of linearly independent pieces.

For $j = 1, 2, \ldots, p$, define $\boldsymbol{\Omega}_j = \{t : \mathbf{f}(t)$ lies on the jth segment$\}$ and $\mathbf{c}_j = \partial\mathbf{f}(t)/\partial t\ |_{t\in\Omega_j}$, which is proportional to the direction of the jth segment. Without loss of generality, assume that t is scaled so that $\partial\mathbf{f}(t)/\partial t\ |_{t\in\Omega_j}$ is constant over each $\boldsymbol{\Omega}_j$. For each fixed t, $\mathbf{f}(t)$ can be expressed as a linear combination of $\{\mathbf{c}_j : j = 1, 2, ..., p\}$ added to $\mathbf{f}(t_0)$, where $\mathbf{f}(t_0)$ is an arbitrary point on $\mathbf{f}(t)$. In other words, $\mathbf{f}(t)$ can be expressed as

$$\mathbf{f}(t) = \mathbf{f}(t_0) + \mathbf{C}\mathbf{v}(t), \quad (7)$$

where $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \ldots \mathbf{c}_p]$, and $\mathbf{v}(t) = [v_1(t)\ v_2(t)\ \ldots\ v_p(t)]^T$ for some set of p random variables $\{v_j(t) : j = 1, 2, \ldots, p\}$ that are functions of t alone. For example, suppose $t_0 \in \boldsymbol{\Omega}_1$, and $\{t_j : j = 1, 2, \ldots, p-1\}$ are defined such that $\mathbf{f}(t_j)$ is the intersection between the jth and the $(j+1)$st segments. Then for each t, $\mathbf{v}(t) = [t_1 - t_0,\ t_2 - t_1,\ \ldots t_{k-1} - t_{k-2},\ t - t_{k-1},\ 0, \ldots, 0]^T$ where k is such that $t \in \boldsymbol{\Omega}_k$.

From Equation (7), it is clear that a piecewise linear curve $\mathbf{f}(t)$ lies in the r-dimensional linear variety $\mathbf{f}(t_0) + \text{span}\ \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_p\}$, where $\mathbf{f}(t_0)$ is an arbitrary point on $\mathbf{f}(t)$ and $r = \text{rank}\ \{\mathbf{C}\} \le p$. Hence, Equations (4)–(6) are applicable to nonlinear variation patterns that can be represented as piecewise linear curves. This has important practical implications with high dimensional manufacturing data, because any nonlinear principal curve can be approximated arbitrarily closely by a piecewise linear one. If only a small (relative to d) number of linear pieces are needed to adequately approximate a principal curve, the preceding results imply that r will be small relative to d and the level of noise reduction achieved by filtering the data will be substantial.

When $\mathbf{f}(t)$ lies entirely in an r-dimensional linear variety, the results of PCA provide an estimate of the linear variety and its dimension. On the other hand, suppose that $\mathbf{f}(t)$ does not lie *entirely* in any low-dimensional linear variety, but that it can be reasonably approximated by a piecewise linear curve. If the approximation is close, then $\mathbf{f}(t)$ will have only a small component that falls outside the linear variety in which the piecewise linear approximation lies, which may be of much lower dimension than d. In this case, the results of PCA will yield an estimate of a linear variety in which $\mathbf{f}(t)$ *approximately* lies. The methods of estimating the number of dominant eigenvalues (i.e., the dimension of the linear variety) discussed in Apley and Shi (2001) are attractive in that the size of an eigenvalue is measured relative to $\sigma_\mathbf{w}^2$. If the size of all components of $\mathbf{f}(t)$ that lie outside a particular linear variety is small relative to $\sigma_\mathbf{w}^2$, it would typically be the case that these components could be neglected.

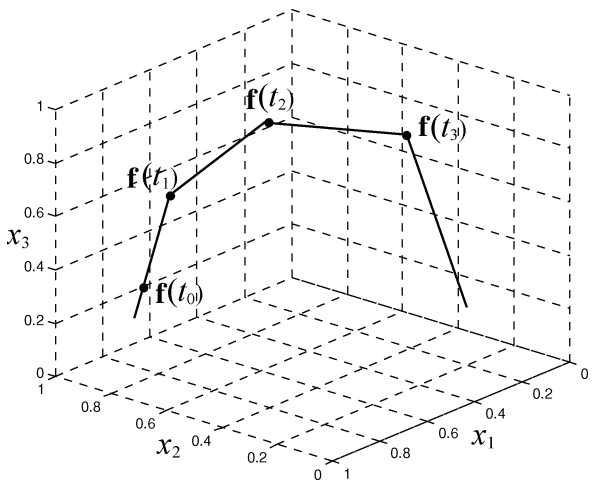

Fig. 3. A piecewise linear principal curve.

5. A PCA-filtered principal curve algorithm

The HS algorithm for estimating a principal curve based on a sample of data $\{\mathbf{x}_i : i = 1, 2, \ldots, N\}$ seeks an estimate $\hat{\mathbf{f}}(\cdot)$ that minimizes $\sum_{i=1}^N \|\mathbf{x}_i - \hat{\mathbf{f}}\|^2$ under certain smoothness constraints. Here, $\|\mathbf{x}_i - \hat{\mathbf{f}}\|^2 = \inf_t \|\mathbf{x}_i - \hat{\mathbf{f}}(t)\|^2$. After conducting PCA on the sample covariance matrix, a set $\{\mathbf{x}_{\mathbf{z},i} : i = 1, 2, \ldots, N\}$ of filtered observations could be generated

in analogy with Equation (6) via $\mathbf{x}_{\mathbf{z},i} = \hat{\mathbf{P}}_{\mathbf{z}}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathbf{x}}) + \hat{\boldsymbol{\mu}}_{\mathbf{x}}$. Here, $\hat{\mathbf{P}}_{\mathbf{z}}$ would be formed from the first $\hat{r}$ eigenvectors $\{\hat{\mathbf{z}}_j : j = 1, 2, \ldots, \hat{r}\}$ of the sample covariance matrix, and $\hat{\boldsymbol{\mu}}_{\mathbf{x}}$ is taken to be the sample average of the observations. One might consider directly applying the principal curve estimation algorithm to the filtered observations by minimizing $\sum_{i=1}^{N} \|\mathbf{x}_{\mathbf{z},i} - \hat{\mathbf{f}}\|^2$. As was discussed in Section 4 and will be demonstrated in Section 7, this reduces the effects of the noise and improves the accuracy of the principal curve estimation.

Aside from the PCA step, the computational expense of this approach would be the same as if the nonlinear variation pattern identification algorithm were applied to the original, unfiltered observations. If $\hat{r}$ is much less than d, however, considerable savings in computational expense can be achieved as follows. Note that since each $\mathbf{x}_{\mathbf{z},i}$ lies in the linear variety $\hat{\boldsymbol{\mu}}_{\mathbf{x}} + \hat{\mathbf{M}}$, where $\hat{\mathbf{M}} = \text{span}\{\hat{\mathbf{z}}_1, \hat{\mathbf{z}}_2, \ldots, \hat{\mathbf{z}}_r\}$, minimizing $\sum_{i=1}^{N} \|\mathbf{x}_{\mathbf{z},i} - \hat{\mathbf{f}}\|^2$ must always result in an $\hat{\mathbf{f}}(\cdot)$ that also lies in $\hat{\boldsymbol{\mu}}_{\mathbf{x}} + \hat{\mathbf{M}}$. Consequently, there is no loss of generality in restricting our search for $\hat{\mathbf{f}}(t)$ to this $\hat{r}$-dimensional linear variety. This can be accomplished by working with the $\hat{r}$-dimensional vectors of PCA scores:

$$\mathbf{y}_i = \hat{\mathbf{Z}}^T(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathbf{x}}), \quad i = 1, 2, \ldots, N, \tag{8}$$

where $\hat{\mathbf{Z}} = [\hat{\mathbf{z}}_1\ \hat{\mathbf{z}}_2\ \ldots \hat{\mathbf{z}}_r]$. Note that $\mathbf{y}_i$ consists of the coefficients of $\mathbf{x}_{\mathbf{z},i} - \hat{\boldsymbol{\mu}}_{\mathbf{x}}$ using $\hat{\mathbf{z}}_j$: $j = 1, 2, \ldots, \hat{r}$ as a basis for the $\hat{r}$-dimensional subspace $\hat{\mathbf{M}}$. Similarly define $\hat{\mathbf{h}}(t) = \hat{\mathbf{Z}}^T(\hat{\mathbf{f}}(t) - \hat{\boldsymbol{\mu}}_{\mathbf{x}})$ to be the coefficients of $\hat{\mathbf{f}}(t) - \hat{\boldsymbol{\mu}}_{\mathbf{x}}$ in $\hat{\mathbf{M}}$. Because we are restricting $\hat{\mathbf{f}}(t) - \hat{\boldsymbol{\mu}}_{\mathbf{x}}$ to lie in $\hat{\mathbf{M}}$, it follows that $\hat{\mathbf{f}}(t) - \hat{\boldsymbol{\mu}}_{\mathbf{x}} = \hat{\mathbf{Z}}\hat{\mathbf{h}}(t)$, or

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{Z}}\hat{\mathbf{h}}(t) + \hat{\boldsymbol{\mu}}_{\mathbf{x}}. \tag{9}$$

Using Equations (8) and (9) and the definition of $\mathbf{X}_{\mathbf{z},i}$, for any t we have

$$\begin{aligned}\mathbf{x}_{\mathbf{z},i} - \hat{\mathbf{f}}(t) &= \hat{\mathbf{Z}}\hat{\mathbf{Z}}^T[\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathbf{x}}] + \hat{\boldsymbol{\mu}}_{\mathbf{x}} - [\hat{\mathbf{Z}}\hat{\mathbf{h}}(t) + \hat{\boldsymbol{\mu}}_x] \\ &= \hat{\mathbf{Z}}[\mathbf{y}_i - \hat{\mathbf{h}}(t)],\end{aligned}$$

so that

$$\begin{aligned}&\|\mathbf{x}_{\mathbf{z},i} - \hat{\mathbf{f}}(t)\|^2 \\ &\quad = \|\hat{\mathbf{Z}}[\mathbf{y}_i - \hat{\mathbf{h}}(t)]\|^2 = [\mathbf{y}_i - \hat{\mathbf{h}}(t)]^t\hat{\mathbf{Z}}^t\hat{\mathbf{Z}}[\mathbf{y}_i - \hat{\mathbf{h}}(t)] \\ &\quad = [\mathbf{y}_i - \hat{\mathbf{h}}(t)]^T[y_i - \hat{\mathbf{h}}(t)] = \|\mathbf{y}_i - \hat{\mathbf{h}}(t)\|^2.\end{aligned}$$

Therefore, choosing a $\hat{\mathbf{f}}(\cdot)$ to minimize $\sum_{i=1}^{N} \|\mathbf{x}_{\mathbf{z},i} - \hat{\mathbf{f}}\|^2$ is equivalent to choosing a $\hat{\mathbf{h}}(\cdot)$ to minimize $\sum_{i=1}^{N} \|\mathbf{y}_i - \hat{\mathbf{h}}\|^2$ and then recovering $\hat{\mathbf{f}}(\cdot)$ from $\hat{\mathbf{h}}(\cdot)$ via Equation (9).

The advantage of working with the $\hat{r}$-dimensional vectors $\{\mathbf{y}_i: i = 1, 2, \ldots, N\}$, as opposed to the d-dimensional vectors $\{\mathbf{x}_{\mathbf{z},i} : i = 1, 2, \ldots, N\}$, is that computational expense and convergence speed are substantially improved.

The PCA-filtered principal curve algorithm is summarized as follows:

Step 1. (Linear variety identification.) Conduct PCA on the sample data and estimate the dimension of the linear variety as the number $\hat{r}$ of dominant eigenvalues.

Step 2. (PCA filtering.) Form $\hat{\mathbf{Z}}$ from the first $\hat{r}$ sample eigenvectors and generate the PCA score vectors $\mathbf{y}_i = \hat{\mathbf{Z}}^T(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathbf{x}})$.

Step 3. (Principal curve estimation.) Use a standard principal curve estimation algorithm (e.g., the HS algorithm) to find the curve $\hat{\mathbf{h}}(\cdot)$ in $\hat{r}$-dimensional space that minimizes $\sum_{i=1}^{N} \|\mathbf{y}_i - \hat{\mathbf{h}}\|^2$. We have used the publicly available software *R*, which can be obtained from http://www.r-project.org.

Step 4. (Principal curve recovery.) Given $\hat{\mathbf{h}}(\cdot)$, recover the d-dimensional principal curve $\hat{f}(\cdot)$ using Equation (9).

We illustrate the variation pattern identification algorithm with the same autobody assembly example introduced in Section 1. An example with much higher dimensional data is considered in the following section. After applying PCA to a sample of $N = 100$ autobodies, it was concluded that there were two dominant eigenvalues. Hence, the principal curve lies approximately in a two-dimensional subspace, which is consistent with the piecewise linear appearance of the variation pattern in Fig. 2. Figure 4(a) is a plot of the eigenvalues. The two corresponding eigenvectors are $\hat{\mathbf{z}}_1 = [0.25\ 0.53\ 0.65\ 0.36\ 0.29\ 0.14]^T$ and $\hat{\mathbf{z}}_2 = [-0.21\ -0.28\ -0.34\ 0.62\ 0.56\ 0.26]^T$. Figure 4(b) is a scatter plot of the two PCA scores y_1 and y_2 for the 100 autobodies, along with the estimated principal curve $\hat{\mathbf{h}}(\cdot)$ in the two-dimensional subspace.

In order to visualize the pattern, we should transform the lower dimensional principal curve $\hat{\mathbf{h}}(\cdot)$ back to the principal curve $\hat{\mathbf{f}}(\cdot)$ in the original six-dimensional data space via Equation (9). The six elements of $\hat{\mathbf{f}}(\cdot)$ (corresponding to the six variables $x_1, \ldots, x_6$) are plotted versus t in Fig. 5. Although the scaling of t is somewhat arbitrary in general, the scaling in Fig. 5 was such that t coincides with the arc length along the principal curve. Compared with the scatter plots in Fig. 2, the plots in Fig. 5 allow a clearer, less noisy, visualization of the nature of the nonlinear pattern.

The ultimate purpose of blindly identifying and visualizing the variation pattern is, of course, to gain insight into the root cause of the variation. By inspection of Fig. 5, certain characteristics of the variation pattern become evident: x_1–x_3, which are all located on the right bodyside, are roughly linearly related. As t varies across its range of values, x_3 varies the most, followed by x_2, and then x_1 varies the least. The features x_4–x_6, which are located on the left bodyside, are not affected by the variation pattern until t approaches its midrange value (i.e., when the movements of x_1–x_3 approach their midrange values). After this point ($t > 1.5$, roughly) the motion of all six features are approximately linearly related, with the features located higher on

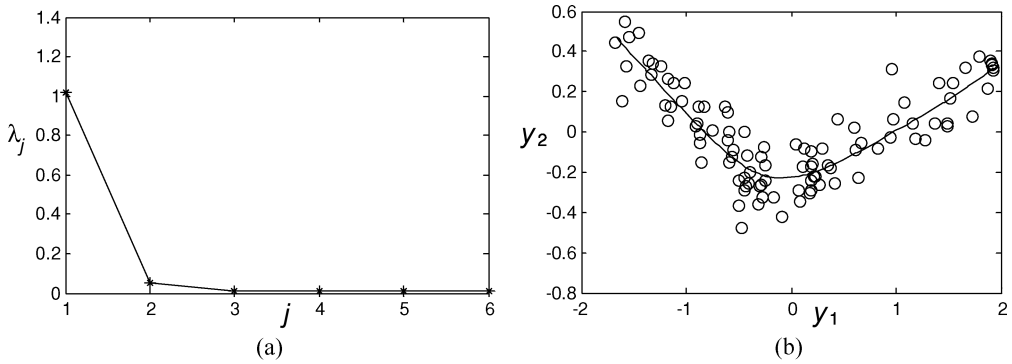

Fig. 4. (a) Eigenvalues and (b) scatter plot of the two dominant PCA scores for the autobody example. Panel (b) also shows the estimated principal curve $\hat{\mathbf{h}}(\cdot)$ in the two-dimensional subspace of dominant PCA scores.

the bodysides (x_3 and x_4) varying to a larger extent. The variation pattern therefore represents a matchboxing of the liftgate opening, in which the left side of the opening begins to matchbox with the right side only after the right side moves through roughly half its total motion. Based on this, a process engineer might speculate that the variation pattern is due to a fixturing problem during the locating of the right bodyside in the framing station (a major assembly station in which the left and right bodysides are joined to the underbody and a set of upper cross-members). When the right bodyside deviates to the right by a large enough amount, it begins to interfere with the upper cross-member. The upper cross-member then interferes with the left bodyside, pulling the left bodyside along with it to the right and causing the matchboxing pattern.

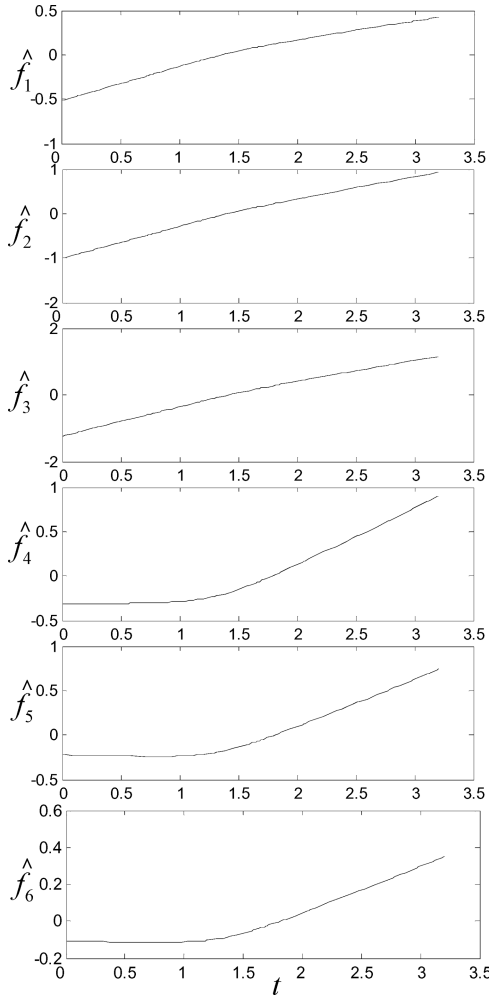

Fig. 5. Plots of the elements of $\hat{\mathbf{f}}(\cdot)$ versus t, illustrating the characteristics of the nonlinear variation pattern.

In this example, we have described how one might interpret the nature of the variation pattern based on inspecting a plot of the elements of $\mathbf{f}(t)$ versus t, such as that shown in Fig. 5. In applications where the measurement data represent dimensional measurements distributed over the surface of a part, visualizing the nature of the variation pattern can be greatly facilitated via the use of interactive graphical animations of the pattern. We describe such an example in the following section.

6. Visualizing high dimensional variation patterns

The purpose of identifying variation patterns is to serve as an aid in identifying and eliminating major root causes of manufacturing variation. In order for a process operator or engineer to effectively interpret the pattern identification results, graphical visualization techniques are critical, especially with high dimensional data. In this section, we illustrate a method for visualizing variation patterns in high dimensional data with an example from sheet metal flanging. In sheet metal panel assembly, the edges of the panels are often formed into a flange in order to add stiffness to the panel or create a mating surface. In shrink flanging, the panel is formed into a concave shape at the same time the flange is formed, which often causes wrinkling in the flange (see Fig. 6). Wrinkling not only mars the appearance of the part, but may also interfere with subsequent sealing and welding operations.

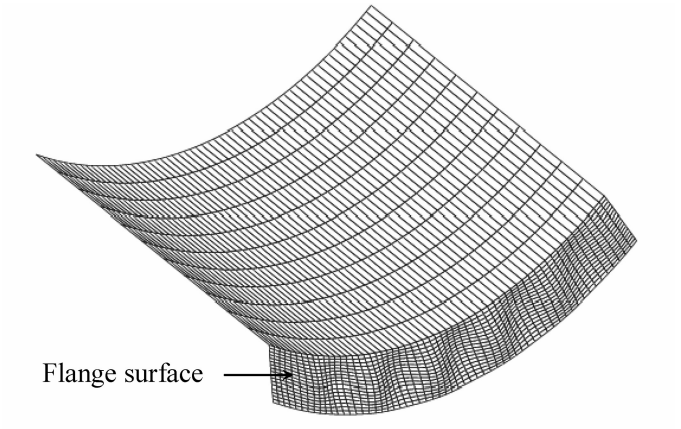

Fig. 6. Illustration of a panel in shrink flanging exhibiting wrinkles on the flange surface.

Wang *et al.* (2001) and the references therein describe analytical and numerical (finite element) models of flange wrinkling. As boundary conditions (e.g., binding force, lubrication, and slippage) change from part to part, the characteristics of the wrinkles change. In the simulation study of this section, the variation pattern represents the position of the wrinkles varying along the length of the flange due to variations in the boundary conditions. The measurement vector $\mathbf{X}$ represents the height of the wrinkles (normal to the surface) over an array of locations distributed across the panel. This could represent point cloud data from a laser scan of the panel, for example. For analysis purposes, we have used $d = 1080$ points distributed uniformly over the panel. The variation pattern was simulated over a sample of $N = 70$ parts, with random noise added, and the results of PCA indicated that there were $\hat{r} = 2$ dominant eigenvalues. Similar to Fig. 4 (a and b), Fig. 7 (a and b) shows the first 11 eigenvalues and a scatter plot of the PCA scores corresponding to the two dominant eigenvalues, as well as the fitted principal curve $\hat{\mathbf{h}}(\cdot)$ in two-dimensional space. The remaining 1069 eigenvalues (not shown in Fig. 7(a)) monotonically decreased to zero. The strong nonlinear (circular) pattern results from what can be viewed as phase shifting as the wrinkles change position along the flange. Although principal curves often have distinct beginning and end points, the underlying (circular) principal curve for this example does not. The reason for this is that the pattern repeats itself as the phase of the wrinkles shifts 360°. Hence, the principal curve is closed for this example, and one might think of joining the beginning and end points of the curve shown in Fig. 7(b).

After transforming $\hat{\mathbf{h}}(\cdot)$ back to the principal curve $\hat{\mathbf{f}}(\cdot)$ in the original 1080-dimensinal space, the nature of the variation pattern could be visualized by graphically illustrating how $\hat{\mathbf{f}}(\cdot)$ varies as t is varied. Figure 8 shows a MATLAB® graphical user interface that was developed for this purpose. The plot of $\hat{\mathbf{f}}(\cdot)$ shown in Fig. 8 changes dynamically as the user moves the slide bar control back and forth. The position of the slide bar represents the value of t for which $\hat{\mathbf{f}}(\cdot)$ is currently plotted. Figure 9 shows five frames (for five different values spanning the range of t) produced from Fig. 8. Only the flange part of the panel is shown. Although the interactive animation would allow one to more clearly visualize the nature of the variation pattern, one can still discern from the five static frames that the pattern represents the wrinkles changing position from part to part.

7. Monte Carlo performance comparison

In this section, Monte Carlo simulation is used to compare the performances of the standard HS algorithm and the PCA-filtered HS algorithm. In all simulations, the data were generated via the model (2). The noise component of $\mathbf{x}$ was

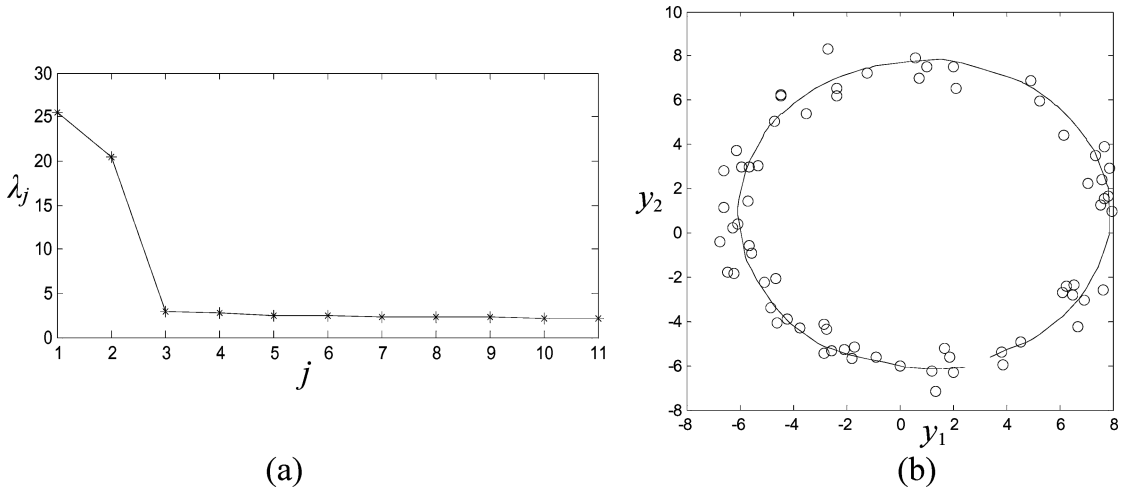

Fig. 7. (a) The first 11 eigenvalues; and (b) the scatter plot of the dominant PCA scores for the simulated shrink flanging example. Panel (b) also shows the estimated principal curve $\hat{\mathbf{h}}(\cdot)$ in the two-dimensional subspace of dominant PCA scores.

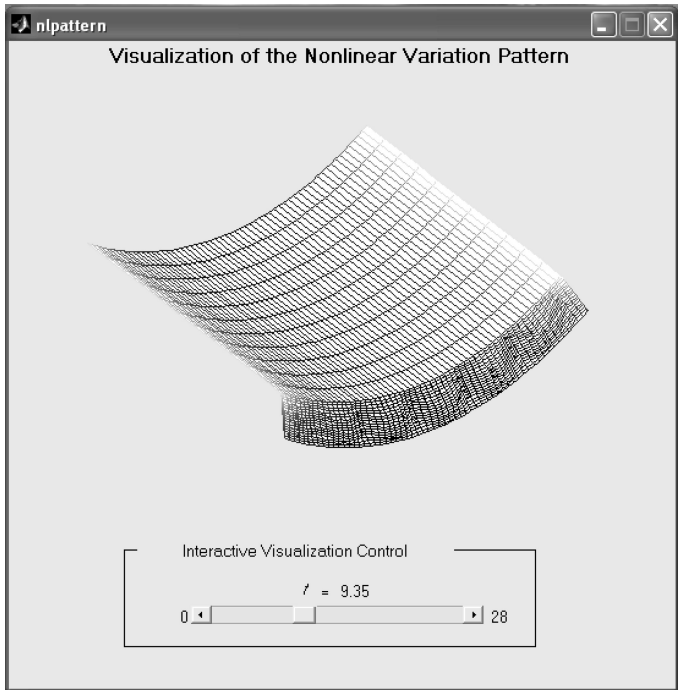

Fig. 8. Graphical user interface for interactively visualizing the nature of a nonlinear variation pattern that represents wrinkling in a shrink flanged panel.

generated from a multivariate Gaussian distribution with covariance matrix $\sigma_{\mathbf{w}}^2\,\mathbf{I}$. The principal curve component of $\mathbf{x}$ was generated as a uniformly distributed point along the specified principal curve $\mathbf{f}(t)$. Denoting the principal curve estimates from the HS algorithm and the PCA-filtered HS

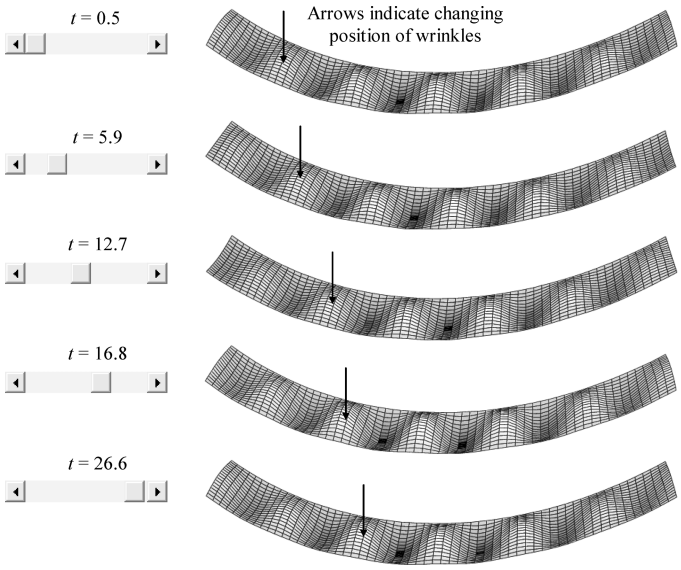

Fig. 9. Illustration of the variation pattern in the flange for five different frames from the interactive visualization interface shown in Fig. 8. The slide bar control in the left panel shows the values of t for the five frames.

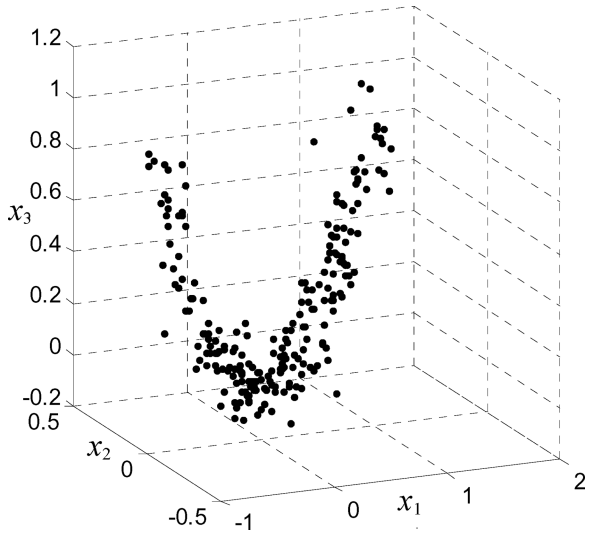

Fig. 10. Illustration of data along a quadratic principal curve with $d = 3$, $r = 2$ and $\sigma_{\mathbf{w}} = 0.1$.

algorithm by $\hat{\mathbf{f}}_{\text{HS}}(\cdot)$ and $\hat{\mathbf{f}}_{\text{HS-PCA}}(\cdot)$, respectively, consider the mean square distances:

$$d_{\text{HS}} = \int \|\hat{\mathbf{f}}_{\text{HS}} - \mathbf{f}(t)\|^2 g(t)\mathrm{d}t, \tag{10}$$

and

$$d_{\text{HS-PCA}} = \int \|\hat{\mathbf{f}}_{\text{HS-PCA}} - \mathbf{f}(t)\|^2 g(t)\mathrm{d}t, \tag{11}$$

as measures of closeness between the estimated and true principal curves. Here, $g(t)$ denotes the probability density of t, which was uniform in all cases. The variable t was scaled so that $||\partial\mathbf{f}(t)/\partial t||$ was constant along the entire length of the curve. In each of the Monte Carlo simulations for the examples discussed below, d_{HS} and $d_{\text{HS-PCA}}$ were averaged over 10 000 Monte Carlo replicates to compare the accuracy of the two principal curve estimation methods.

A quadratic principal curve lying in a two-dimensional linear variety of $\Re^d$ was used for all simulations. The principal curve was constructed by first generating the curve $x_2 = x_1^2$ for $x_1 \in [-1, 1]$, which is illustrated in Fig. 10 for $d = 3$. The curve was then multiplied by a Householder mirror reflection matrix in d-dimensional space to transform it without changing its shape or size so that it still fell in a linear variety of dimension $r = 2$.

Table 1 compares the performance of the HS algorithm and the PCA-filtered HS algorithm for various d and $\sigma_{\mathbf{w}}$, with a sample size of $N = 200$. Note that the value of σ_{w} in Fig. 10 corresponds to the midrange value of 0.1 in Table 1. The extent to which PCA filtering improves the accuracy (as measured by $d_{\text{HS}}/d_{\text{HS-PCA}}$) varied between 1.48 and 6.96, with larger relative improvement for larger d (higher dimensional data) and/or large $\sigma_{\mathbf{w}}$ (larger noise-to-signal ratio). Simulations were also run for $N = 100$ and $N = 400$. The individual performance measures d_{HS} and $d_{\text{HS-PCA}}$ were

Table 1. Performance comparison of the HS and PCA-filtered HS algorithms for quadratic principal curves ($N = 200$)

d	σ_w	d_{HS}	$d_{HS\text{-}PCA}$	d_{HS}/d_{HS-PCA}
8	0.05	0.0031	0.0021	1.48
8	0.1	0.021	0.0089	2.36
8	0.2	0.43	0.16	2.69
16	0.05	0.0034	0.0019	1.79
16	0.1	0.056	0.021	2.67
16	0.2	0.63	0.19	3.32
64	0.05	0.0065	0.0025	2.60
64	0.1	0.24	0.059	4.07
64	0.2	3.56	0.68	5.24
100	0.05	0.022	0.0052	4.23
100	0.1	0.56	0.098	5.71
100	0.2	5.64	0.81	6.96

roughly inversely proportional to N, and their ratio was virtually the same for all values of N.

8. Conclusions

This paper develops a methodology for representing and blindly identifying nonlinear variation patterns in high dimensional manufacturing measurement data. The nonlinear variation pattern model is an extension of previous linear models that provides more accurate representation of complex nonlinear phenomena. To overcome computational and accuracy problems caused by high dimensionality, thereby making the approach more suitable for large-scale manufacturing databases, we used a common PCA-based dimensionality reduction step prior to principal curve estimation. We showed that the dimensionality reduction step sacrifices very little information with the type of variation patterns that one would typically encounter in manufacturing. The Monte Carlo simulation results demonstrate that the PCA-filtering can also substantially improve the accuracy of the variation pattern estimation.

The paper also describes the use of graphical visualization methods for aiding in interpreting the blindly identified variation patterns. This allows potentially valuable information to be extracted from large volumes of manufacturing measurement data. Together, the pattern identification and visualization techniques provide an effective tool for diagnosing major root causes of manufacturing variation.

As a final comment, we point out that there are many nonlinear pattern identification and other data mining methods (e.g., those discussed in Hastie *et al.* (2002)) that one might have considered for this problem. The body of candidate methods shrinks considerably when one restricts attention to only those that fall into the category of unsupervised learning, as is required in our application. Furthermore, we desired an approach that could identify the nature of the variation pattern as blindly and generically as possible, without assuming any particular structural form for the functional relationship between the different variables (e.g., radial basis functions, sigmoidal functions associated with neural networks, wavelet basis functions, etc.). With these requirements, principal curve estimation (or the self-organizing maps of Kohonen (Kohonen, 1990), which may be viewed as a discrete version of principal curves (Cherkassky and Mulier, 1998)) is perhaps the most suitable. Another attractive aspect of the principal curve approach is that it provides a convenient means for visualizing the variation pattern (refer to Figs. 8 and 9) following its blind identification, thereby helping process engineers to discover the nature and root cause of the variation.

References

Apley, D.W. and Lee, H.Y. (2003) Identifying spatial variation patterns in multivariate manufacturing processes: a blind separation approach. *Technometrics*, **45**, 220–235.

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**, 1–12.

Banfield, J.D. and Raftery, A.E. (1992) Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. *Journal of the American Statistical Association*, **87**, 7–16.

Barton, R.R. and Gonzalez-Barreto, D.R. (1996) Process-oriented basis representation for multivariate process diagnostics. *Quality Engineering*, **9**, 107–118.

Ceglarek, D. and Shi, J. (1996) Fixture failure diagnosis for the autobody assembly using pattern recognition. *ASME Journal of Engineering for Industry*, **118**, 55–66.

Chang, K. and Ghosh, J. (2001) A unified model for probabilistic principal surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **23**, 22–41.

Cherkassky, V. and Mulier, F. (1998) *Learning From Data: Concepts, Theory, and Methods*, Wiley, New York, NY.

Delicado, P. (2001) Another look at principal curves and surfaces. *Journal of Multivariate Analysis*, **77**, 84–116.

Delicado, P. and Huerta, M. (2003) Principal curves of oriented points: theoretical and computational improvements. *Computional Statistics*, **18**, 293—315.

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

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

Hastie, T., Tibshirani, R. and Friedman, J. (2002) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, NY.

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. (2002) *Applied Multivariate Statistical Analysis*, 5th edn., Prentice Hall, Upper Saddle River, NJ.

Kégl, B., Krzyzak, A., Linder, T. and Zeger, K. (2000) Learning and design of principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**, 281–297.

Kohonen, T. (1990) The self-organizing map, *Proceedings of the IEEE*, **78**, 1464–1479.

Lee, H.Y. and Apley, D.W. (2004) Diagnosing manufacturing variation using second-order and fourth-order statistics. *International Journal of Flexible Manufacturing Systems*, **16**, 45–64.

Tibshirani, R. (1996) *Principal Curves and Neural Networks*, Cambridge University Press, Cambridge, UK.

Wang, X., Cao, J. and Li, M. (2001) Wrinkling analysis in shrink flanging. *ASME Journal of Manufacturing Science and Engineering*, **124**, 426–432.

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

Appendix

Proofs of Equations (4) and (5). First note that because $\mathbf{f}(t) \in \mathbf{a}_0 + \mathbf{M}$ for all t, it must be the case that its mean $\boldsymbol{\mu}_\mathbf{f}$ also lies in $\mathbf{a}_0 + \mathbf{M}$. Also, because $\mathbf{W}$ is zero-mean, $\boldsymbol{\mu}_\mathbf{x} = \boldsymbol{\mu}_\mathbf{f}$. Hence, $\boldsymbol{\mu}_x \in \mathbf{a}_0 + \mathbf{M}$, so that the linear variety $\mathbf{a}_0 + \mathbf{M}$ is the same as $\boldsymbol{\mu}_\mathbf{x} + \mathbf{M}$. By the independence of $\mathbf{W}$ and t:

$$\begin{aligned}\Sigma_\mathbf{x} &= E[(\mathbf{X} - \boldsymbol{\mu}_\mathbf{x})(\mathbf{X} - \boldsymbol{\mu}_\mathbf{x})^T] = E[(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x} + \mathbf{W})(\mathbf{f}(t) \\ &\quad -\boldsymbol{\mu}_\mathbf{x} + \mathbf{W})^T] = E[(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})^T] + \sigma_\mathbf{w}^2\mathbf{I}.\end{aligned}$$

Now consider any unit-norm vector $\mathbf{z} \in \mathbf{M}^\perp$ (the orthogonal complement of $\mathbf{M}$). Because $\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x}$ lies in $\mathbf{M}$ and $\mathbf{z}$ lies in the orthogonal complement of $\mathbf{M}$, $(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})^T\mathbf{z} = 0$ for all t. Therefore, $\Sigma_\mathbf{x}\mathbf{z} = E[(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})^T\mathbf{z}] + \sigma_\mathbf{w}^2\mathbf{I}\mathbf{z} = \sigma_\mathbf{w}^2\mathbf{z}$. Hence, $\mathbf{z}$ is an eigenvector of $\Sigma_\mathbf{x}$ with eigenvalue $\sigma_\mathbf{w}^2$. Because $\mathbf{M}^\perp$ is a $d - r$ dimensional subspace, there exists an orthonormal set of $d - r$ such eigenvectors, which we denote $\{\mathbf{z}_{r+1}, \mathbf{z}_{r+2}, \ldots, \mathbf{z}_d\}$. Because the remaining eigenvectors $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_r\}$ are orthogonal to $\{\mathbf{z}_{r+1}, \mathbf{z}_{r+2}, \ldots, \mathbf{z}_d\}$, they must all lie in the r-dimensional subspace $\mathbf{M}$. Consequently, $\mathbf{M} = \text{span}\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_r\}$, which completes the proof of Equation (5).

Because $\mathbf{z}_j \in \mathbf{M}$ for $1 \le j \le r$, its corresponding eigenvalue is

$$\begin{aligned}\lambda_j &= \mathbf{z}_j^T(\lambda_j\mathbf{z}_j) = \mathbf{z}_j^T(\boldsymbol{\Sigma}_\mathbf{x}\mathbf{z}_j) \\ &= \mathbf{z}_j^T\left\{E[(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})^T] + \sigma_\mathbf{w}^2\mathbf{I}\right\}\mathbf{z}_j \\ &= E\left[\mathbf{z}_j^T(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})^T\mathbf{z}_j\right] + \sigma_w^2 \\ &= E\left[\mathbf{z}_j^T(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})^2\right] + \sigma_\mathbf{w}^2 > \sigma_\mathbf{w}^2.\end{aligned}$$

That $E\left[(\mathbf{z}_j^T(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x}))^2\right]$ is strictly greater than zero follows from the condition that no other linear variety in which $\mathbf{f}(t)$ lies has dimension smaller than r. Indeed, if $\mathrm{E}\left[(\mathbf{z}_j^T(\mathbf{f}(t) - \boldsymbol{\mu}_\mathbf{x})^2)\right] = 0$ for some $1 \le j \le r$, this would imply that $\mathbf{f}(t)$ lies in the $r - 1$ dimensional linear variety $\boldsymbol{\mu}_\mathbf{x} + \text{span}\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{j-1}, \mathbf{z}_{j+1}, \ldots, \mathbf{z}_r\}$. ■

Biographies

Daniel W. Apley received BS and MS degrees in Mechanical Engineering, an MS degree in Electrical Engineering, and a PhD degree in Mechanical Engineering in 1990, 1992, 1995, and 1997, respectively, all from the University of Michigan. He was an Assistant Professor of Industrial Engineering at Texas A&M University from 1998 to 2003. Since September, 2003, he has been an Associate Professor of Industrial Engineering and Management Sciences and the Morris E. Fine Professor of Manufacturing at Northwestern University. His research interests include manufacturing variation reduction, with particular emphasis on processes in which advanced measurement, data collection, and automatic control technologies are prevalent. Much of his work involves the use of engineering models, knowledge and information within a statistical framework for analyzing large sets of process and product measurement data. He was an AT&T Bell Labs PhD Fellow in Manufacturing Science from 1993–1997, and he received the NSF Career Award from 2001–2006 for his research and teaching. He currently serves on the Editorial Board of the *Journal of Quality Technology* and as an Associate Editor of *Technometrics*. He is a member of IEEE, ASME, IIE, SME, and INFORMS.

Feng Zhang received B.S. and M.S. degrees in Control Theory and Engineering from Tsinghua University, China, in 1996 and 1999, respectively, and a Ph.D degree in Industrial and System Engineering from Texas A&M University, College Station, TX, in 2004. He is currently an Operations Research Analyst at Fairchild Semiconductor, South Portland, ME. His research encompasses many aspects of data mining and statistical quality control, which involves the integration of pattern recognition and visualization with multivariate data analysis for identifying systematic causes of manufacturing variation. He is a member of INFORMS, IEEE, and DSI.