A Comparative Study on the Classification of Engineering Surfaces
with Dimension Reduction and Coefficient Shrinkage Methods

Min Zhang               Ji Zhu
Department of Mechanical Engineering    Department of Statistics
The University of Michigan    The University of Michigan
Ann Arbor, MI 48109-2136    Ann Arbor, MI 48109-2136
Email: mzhangz@umich.edu    Email : jizhu@umich.edu
Phone: 1 734-764-5391        Phone : 1 734-936-2577
Fax : 1 734-936-0363

Dragan Djurdjanovic     Jun Ni
Department of Mechanical Engineering   Department of Mechanical Engineering
The University of Michigan                          The University of Michigan
Ann Arbor, MI 48109-2136                            Ann Arbor, MI 48109-2136
Email: ddjurdja@umich.edu                          Email: junni@umich.edu
Phone: 1 734-763-9975                              Phone: 1 734-936-2918
Fax : 1 734-936-0363                   Fax : 1 734-936-0363

Keywords: surface classification, dimension reduction, coefficient shrinkage, comparative study.
Abstract:

As more automated and accurate surface inspection devices enter the manufacturing process, engineers collect a larger amount of surface inspection data, in terms of storage space and the number of parameters to characterize the surface, but sometimes smaller in terms of the number of coherent surface observations. In these cases, more features are preferable to characterize engineering surfaces for capturing the details of the surface finish patterns. When the number of surface parameters exceeds the number of collected surface observations, a difficulty with the dimensionality emerges in classification. This paper has researched the accuracy and interpretability of using the dimension reduction and coefficient shrinkage methods in combination with the logistic model to deal with this dimensionality problem in engineering surface classification. Five methods for dimension reduction and coefficient shrinkage are selected and compared. These are subset selection (Sub), principal component analysis (PCA), partial least squares (PLS), ridge regression (Ridge) and Lasso. A case study is used to illustrate their effectiveness by classifying 30 pump body surfaces with 40 surface feature parameters. The obtained results show that the dimension reduction methods, the PCA and the PLS, could achieve higher classification accuracies but their results are not interpretable. The Sub could achieve higher accuracy in this case, but the discrete parameter selection process is aggressive. Finally, the coefficient shrinkage methods, the Ridge and Lasso, their classification results are interpretable for process faults diagnosis purposes. However, the accuracies are lower than the other methods.
1 Introduction

Quantitative measures are obtained from inspection data of engineering surfaces. Classification and pattern recognition of these surface measures can assist in machine condition monitoring and failure analysis of engineering components. Conventionally, a small subset of the numerical measures called surface features is first selected to characterize the surface finish pattern. Then, enough surface observations are often collected to classify the engineering surfaces [1-2]. As more automated and accurate surface inspection devices enter the manufacturing process, engineers collect a larger amount of surface inspection data, in terms of storage space and the number of parameters to characterize the surface, but sometimes smaller in terms of the number of coherent surface observations. Examples can be found when an engineering surface of interest is very large and complicated, which is often the case with automotive transmission valve bodies, automobile pump bodies, etc. In these cases, more features are preferable to characterize engineering surfaces for capturing the details of the surface finish patterns. In addition, when the classification is used for diagnostic purposes, the use of the so-called local surface parameters [3], which are the surface parameters summarized from sub-areas by dividing the entire original surface into smaller regions, could assist in identifying the location of error sources, e.g. fixture setup differences for surface distortion problems. However, in the case when the number of surface parameters exceeds the number of collected surface observations, a difficulty with the dimensionality emerges in classification: the degrees of freedom are insufficient to accurately classify the surface observations. Namely, if the number of surface feature parameters \( p \) increases beyond the number of the observed surface samples \( n \), the accuracy of
classification will deteriorate and meaningless classification results may be generated [4]. To overcome this problem, researchers tend to use the non-parametric classification methods. For example, Podsiadlo and Stachowiak [5] proposed a novel hybrid fractal-wavelet dissimilarity measure method to bypass this dimensionality problem, which belongs to one of the non-parametric classification algorithms. However, the weakness to use the non-parametric method for surface classification lies in two aspects. One is that given a surface sample, only the surface class label is determined but no estimate of the underlying probability of the classification is provided. Another one is that the non-parametric method is not able to identify the most significant surface parameters from the classification results for diagnostic purposes.

In this paper, the parametric methods are used to classify large engineering surfaces mainly for diagnostic purposes. Two simple but powerful parametric methods that can both estimate the probability and identify significant surface parameters are linear discriminant analysis and logistic regression. From Hastie [6], logistic regression is concluded as a safer, more robust method than the linear discriminant analysis, relying on fewer assumptions. So the logistic regression will be used in this study. Since the ordinary logistic regression (OLS) by itself is not able to address the dimensionality problem, while a number of approaches referred to as dimension reduction and coefficient shrinkage methods exist in the applied statistical literature, dealing with the search for a low-dimensional manifold that embeds a set of high-dimensional data [6], the combination of the logistic model and the dimension reduction and coefficient shrinkage methods would have the potential to solve the dimensionality problem, provide estimates of probabilities and identify significant surface parameters for diagnostic purposes.
In this paper, the classification of engineering surfaces with a logistic regression model in conjunction with the dimension reduction and coefficient shrinkage methods to tackle with the dimensionality problem \( n < p \) will be focused. Five commonly used dimension reduction and coefficient shrinkage approaches from applied statistics literature are selected for a comparative study. Figure 1 shows the relationship between these methods, which are: 1) subset selection (Sub) [7]; 2) principal component analysis (PCA) [6]; 3) partial least squares (PLS) [8]; 4) ridge regression (Ridge) [9]; and 5) least absolute and shrinkage and selection operator (Lasso) [10]. These five methods can be divided into two categories. One is called dimension reduction techniques, which are utilized to achieve the mapping of a multidimensional space into a space with fewer dimensions. PCA is the simplest method for dimension reduction. PLS is similar to PCA but involves more information in constructing the model. Both of them are defined as the “methods using derived input directions” in [6]. They improve the classification accuracy in case of dimensionality. However, their classification results are not interpretable since the reduction of the dimensionality is based on the linear/non-linear combination of the original inputs. In order to provide interpretability, another category of methods are also chosen in this study, which are called variable selection/coefficient shrinkage methods. Subset selection is the simplest way to reduce dimensionality and produce an interpretable model with possibly lower classification errors. However, it’s a discrete variable selection process and only part of the surface feature parameters will be chosen for classification model while the rest discarded. In this case, subset selection often exhibits high variance [7]. Two more complicated methods, Ridge and Lasso called coefficient shrinkage methods, are also included in this study. Ridge continuously shrinks
the coefficients to reduce dimensionality and Lasso finally shrinks the insignificant coefficients to zeros and achieves the selection via shrinkage. Compared with Sub, they are more continuous and do not suffer as much from high variability in order to provide interpretable results. Since both of them use penalties on the sizes of the surface feature parameters to achieve shrinkage, they are known as the “penalized methods”.

![Dimension Reduction and Coefficient Shrinkage Methods](image)

*Figure 1 Dimension Reduction and Coefficient Shrinkage Methods*

The remainder of this paper is organized as follows: section 2 briefly reviews the elemental procedures of classification by ordinary logistic regression. Section 3 contains mathematical descriptions of the selected five dimension reduction and coefficient shrinkage approaches to deal with the dimensionality problem. A case study to classify two sets of pump body surfaces (feature parameters $p = 40$; surface sample size $n = 30$) with the logistic model in combination with the dimension reduction and coefficient
shrinkage methods will be presented in section 4. Finally, section 5 offers conclusions of this comparative study and several guidelines for possible future work.

2 Logistic Regression for Classification

In a standard two-class classification problem, one is given a set of training data 
\((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\) , where the input \(x\) is a \(p\) dimensional vector \(x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T\), the output \(y_i\) is qualitative and assumes binary values \(\{0,1\}\). In the case of surface classification, the input \(x\) is a set of surface feature parameters while the output \(y\) could be the class labels of two predefined surface categories with engineering knowledge. The logistic regression model arises from the desire to model the conditional probabilities of the two classes via linear functions in \(x\), while at the same time ensuring that all the conditional probabilities sum up to one. Assume the training data is independently and identically distributed from an probability distribution \(P(X,Y)\), the logistic regression model has the form

\[
p_{y=1}(x) = P(Y = 1 \mid X = x) = \frac{\exp(\beta_0 + \beta^T x)}{1 + \exp(\beta_0 + \beta^T x)}
\]

(1)

\[
p_{y=0}(x) = P(Y = 0 \mid X = x) = \frac{1}{1 + \exp(\beta_0 + \beta^T x)}
\]

(2)

where \(p_{y}(x) = P(Y = y \mid X = x)\) is the conditional probability of a surface being in class \(y\) given its surface feature parameters obtained from \(X = x\), and \(\beta = (\beta_0, \beta)^T\) is a set of regression coefficients, which are model parameters that can often be fit by maximum likelihood (ML) method [11]. Given \(n\) training samples with known class labels
(x_1, y_1), (x_2, y_2), ..., (x_n, y_n), the multinomial log-likelihood in the two-class case can be written as the joint probability of the n observations

\[
\log(L(\hat{\beta})) = \sum_{i=1}^{n} \left\{ y_i \cdot \log(p_{y_i}(x_i)) + (1 - y_i) \cdot \log(1 - p_{y_i}(x_i)) \right\}
\]

\[
= \sum_{i=1}^{n} \left\{ y_i \cdot (\beta_0 + \beta^T x_i) - \log(1 + e^{\beta_0 + \beta^T x_i}) \right\}
\]  \hspace{1cm} (3)

In order to maximize the log-likelihood and obtain the regression coefficients \( \hat{\beta} \) for which the observed samples are most likely to have been generated, the derivative of the log-likelihood is set to zero. This results in \( p + 1 \) nonlinear equations in \( \hat{\beta} \), which are typically solved with the iteratively reweighted least squares (IRWLS) algorithm till its convergence. More details about the fitting procedure can be found in [6]. With the fitted coefficients from n training samples, the logistic model is then able to classify new samples with surface feature parameters \( x_{\text{new}} \) to one of the pre-defined classes with a probability \( p_{y_j}(x_{\text{new}}) \).

Typically, before the fitted model is used to classify new surfaces, another set of surface samples called a validation dataset is necessary to assess the classification performance of the fitted model. Since often one does not have the luxury of having these additional samples, cross validation techniques are suitable to evaluate the classification accuracy before the model is utilized for new samples. There is extensive literature documenting a variety of cross validation methods, e.g. K-fold cross-validation, leave-V-out, bootstrap, etc.

For the K-fold cross-validation (CV) [6], the n training samples are divided into K subsets. For the \( k^{th}, k = 1, 2, ..., K \) subset, misclassification error is calculated from the
model fitted using of the other $K - 1$ subsets, and finally the misclassification errors are averaged for $k = 1, 2, \ldots, K$. The so-called cross-validated misclassification error for the fitted model is then expressed as

$$CV \text{ misclassification error} = \frac{1}{K} \sum_{k=1}^{K} \frac{\# \text{ of samples misclassified in } k^{th} \text{ part}}{\text{total } \# \text{ of samples in } k^{th} \text{ part}}$$  \hspace{1cm} (4).$$

There are other validation methods in literature, e.g. leave-$V$-out, bootstrap, etc. Compared with the $K$-fold cross-validation, in which only the omitted subsets are used to compute the misclassification error, the leave-$V$-out is elaborate and expensive because it involves leaving all possible subsets of $V$ cases. Bootstrap seems to work better than the $K$-fold cross-validation. Instead of repeatedly analyzing subsets of the surface data, it repeated analyze sub-samples of the data. However, it usually requires a pool of 50-2000 sub-samples to repeat the analysis. Considering the application in this study that engineering surfaces with relative small surface samples will be used, the $K$-fold cross-validation is selected for model validation.

3 Dimension Reduction and Coefficient Shrinkage Methods

The ordinary logistic regression described in Section 2 has been successfully and widely applied in engineering surface classification, when only a few parameters are used to characterize the surfaces [12]. However, its implementation raises several concerns when the surface observations are less than the surface feature parameters ($n < p$): (1) unnecessary surface feature parameters add noise and the degree of freedom is wasted; (2) the classification accuracy deteriorates, if collinearity exists
between surface feature parameters. To overcome these problems, the dimension reduction and coefficient shrinkage methods are introduced. In this section, the mathematical descriptions of five selected dimension reduction and coefficient shrinkage methods mentioned in Section 1 are provided.

### 3.1 Subset Selection (Sub)

Subset selection denotes a procedure to retain a subset of the significant surface feature parameters while eliminating the insignificant ones in the ordinary logistic regression model. Heuristically, three procedures are all able to evaluate the significance of feature parameters: backward elimination, forward selection, and stepwise regression [7]. Backward elimination starts with the full model (including all the potential surface parameters) and eliminates the insignificant ones sequentially. Forward selection starts with a null model (no parameters) and then adds the surface features which best fit the model. The stepwise regression is the combination of both forward selection and backward elimination. The selection or elimination strategy in all three procedures is made on some criterion such as the $F$ statistic, adjusted $R^2$, etc. Details of the criterion could be found in [7]. One obvious disadvantage of subset selection methods is that they tend to be aggressive and pick models smaller than desirable for classification purposes. Furthermore, in the cases $n < p$, backward elimination is not applicable.

### 3.2 Principal Component Analysis (PCA)

The basic idea of the principal component analysis is to construct a series of linear combinations of the original surface feature parameters and then use them as the
input to the ordinary logistic regression. Define a $n \times p$ matrix $X$ containing all the surface feature parameters of $n$ observations. The linear combination $z_m = X \cdot u_m$ referred to as the principal component is the projection of all the surface parameters on a small number of principle directions $u_m, m = 1, 2, ..., M$. The construction of the principle directions maximizes the variance of the principal components

$$\arg \max_{u_m, u_m = 1} Var(X \cdot u_m)$$

in which the condition $u_m^T u_l, l = 1, 2, ..., m - 1$ ensures that one direction $u_m$ is orthogonal to all other previous directions. Once the principle directions $u_m$ and principle components $z_m$ are obtained, ordinary logistic regression is performed by regressing $y$ on the significant principal components $z = (z_1, z_2, ..., z_M)^T$. With the IRWLS algorithm mentioned in Section 2, the regression coefficient between $y$ and $z$ are computed as $\tilde{\gamma}$, after which the classification model is completed by calculating the coefficients between the original $y$ and $x$ as $\tilde{\beta} = (u_1, u_2, ..., u_M)^T \cdot \tilde{\gamma}$. In this paper, the optimal selection of the significant number of the principal components $M$ is conducted via the $K$-fold cross-validation of the training samples to minimize the CV misclassification errors.

### 3.3 Partial Least Squares (PLS)

The technique of partial least squares is similar to the PCA. It also produces linear combinations of the original surface parameters, so that no correlation exists between the new parameters used in the classification model. However, PLS and PCA
differ in the way how they extract the principal directions. PCA ignores the information in $Y, Y = (y_1, y_2, \ldots, y_n)^T$ when building the principal components $z_m, m = 1, 2, \ldots, M$, and PLS produces the directions reflecting the relationship between $Y$ and $X$. By constructing a score matrix $T = X \cdot W$, where $W$ is called a weight matrix, each column of $W$ is obtained by sequentially solving

$$\arg \max_{w_m \neq w_j = 1} \text{Cov}(Xw_m, Y). \quad (6)$$

Ordinary logistic regression procedure is performed by regressing $Y$ on $T$ such that the regression coefficient $\tilde{\lambda}$ is calculated with ML method and the IRWLS algorithm. Finally, the coefficient between $Y$ and $X$ is computed as $\tilde{\beta} = W \cdot \tilde{\lambda}$. More details about the procedure for PLS can be referred to [8].

### 3.4 Ridge Regression (Ridge)

As mentioned in Section 1, ridge regression belongs to the penalized methods to deal with dimensionality, which shrink the regression coefficients by imposing a penalty on their size. Ridge coefficients, first proposed by Hoerl and Kennard in [9, 13], minimize a penalized residual sum of squares in ordinary linear regression models. For generalized linear models such as logistic regression, penalized maximum likelihood estimation has been proposed more recently by Harrell et al. [14]. In the case of two-class classification with logistic model, a constraint is added to control the spread of the coefficients while maximizing the log-likelihood of the logistic model

$$\arg \max_{\sum |\beta| \leq \delta} \left\{ \sum_{i=1}^{n} \left[ y_i \cdot (\beta_0 + \beta^T x_i) - \log(1 + e^{\beta_0 + \beta^T x_i}) \right] \right\}. \quad (7)$$
Here $t \geq 0$ is the complexity parameter that controls the amount of shrinkage of the coefficients. Since the regression coefficients are continuously shrinking during the fitting process with the tuning of parameter $t$, ridge regression is also referred to as “shrinkage during fitting”. The optimal selection of the parameter $t$ is determined by the $K$-fold cross-validation of the training samples to minimize the CV misclassification errors.

3.5 Least Absolute and Shrinkage and Selection Operator (Lasso)

As described in Section 3.1 and 3.4, subset selection provides interpretable models but can result in high variance since it is a discrete process, while ridge regression is a continuous process but it does not set any of the coefficients to 0 and will not give an easily interpretable model. In attempt to retain the good features of both subset selection and ridge regression to deal with dimensionality, a technique called least absolute shrinkage and selection operator was proposed by Tibshirani [10] to shrink some coefficients and set others to exact zero by imposing the constraint

$$\sum_{j=1}^{p} |\beta_j| \leq s.$$  

By solving

$$\arg\max_{\sum_{j=1}^{p} |\beta_j| \leq s} \left\{ \sum_{i=1}^{n} \left[ y_i \cdot (\beta_0 + \beta^T x_i) - \log(1 + e^{\beta_0 + \beta^T x_i}) \right]\right\}$$

the Lasso shrinks the coefficients of the logistic model toward zero. While the penalty is added on the absolute values of the regression coefficients, some of them are potentially shrunk to zero. Since the surface feature parameters with zero coefficients can be omitted, selection of surface parameters is achieved with the implementation of Lasso. For this reason, the Lasso may be referred to as “selection through shrinkage”. Similar
to Ridge, the tuning parameter $s$ is selected via the $K$-fold cross-validation of the training samples to minimize the CV misclassification errors. More details about Lasso can be found in [10, 15].

4 Case Study

In this section, a case study is presented to demonstrate how the logistic model with dimension reduction and coefficient shrinkage methods could improve the classification accuracy and interpretability in case of dimensionality. All the surface measurements involved in the case study were collected from a Coherix ShaPix device [16]. The optical system of this non-contact surface inspection machine is schematically shown in Figure 2. The coherent light produced from a multiple-wavelength laser system is split by a beam splitter into two parts, one of which illuminates an object from top down, while the other illuminates a reference surface. Light scattered from the reference surface and the object is re-combined at the beam splitter and a phase image is recorded on the CCD array detector. The phase-shifting and multiple wavelength tuning technologies are used to achieve high accuracy ($\pm 1 \ \mu m$) for a large measurement range (centimeters level) on surfaces of different materials, such as optical mirror, machined metal, and plastic. One million points are measured in about 1 minute cycle time to deliver a high definition description of the object surface profile in the 300 mm x 300 mm field of view, which indicates the lateral resolution of this measurement device is around 0.3 mm.
The engineering surfaces selected for this case study are the inner flat surfaces of the pump bodies machined by a major domestic car manufacturer. Two sets of data have been collected with 15 pump bodies finished from machine A forming Set A of surfaces, and another 15 pump bodies from machine B forming Set B of surfaces. Two samples of the color-coded measurement results from each dataset are shown in Figure 3. It has been observed by the quality engineers in the field that parts in set A have less leakage problems than the ones in set B, which may be resulted from the different levels of surface distortion of the pump body surfaces. Engineers suspect that the root causes of the surface problems may come from the fixture setup of each machine.
Figure 3 Surface Measurement Results of Two Part Samples from Dataset A & B Respectively.

Since the total surface observations are $n=30$ in this case, a larger number of surface parameters to characterize the pump body surfaces ($p>30$) will be necessary to illustrate the effectiveness of dimension reduction and coefficient shrinkage techniques in classification. In order to obtain $p>30$ features, each of the original measured surface is divided into $p$ smaller regions. A so-called local flatness parameter [3] is calculated within each smaller region, which is similar to the total roughness $R_t$ parameter in Geometric Dimensioning and Tolerancing (GD&T) convention. More specifically, the local flatness parameter within each region $LF_p$ is defined as the vertical distance between two planes parallel to the reference plane and enveloping the whole acquired surface. The obtained $p$ ($p>n$) features are then used to characterize the engineering surfaces for classicization. In this case study, $p=40$ is selected just for convenience. Forty markers are used to divide each measured surface into grids. Locations of the 40 markers in this study are shown in Figure 4. A local flatness parameter around each
marker is calculated using the surface height value within a circular area of 80 pixels (around 24mm) in diameter similar to the way of calculating the total roughness parameter $R_t$.

![Diagram of pump surfaces]

**Figure 4 Local flatness descriptions of the pump surfaces**

With the definition of the local flatness parameter, the surface feature for the $i^{th}, i = 1, 2, ..., n = 30$ observation is denoted as $x_i = (LF_{i,1}, LF_{i,2}, ..., LF_{i,40})^T$. The corresponding output variable is defined as

$$y_i = \begin{cases} 0 & \text{if surface is from set } A \\ 1 & \text{if surface is from set } B \end{cases}; \quad i = 1, 2, ..., n = 30$$

The total 30 pump body samples $(x_1, y_1), (x_2, y_2), ..., (x_{30}, y_{30})$ are divided into a training set (20 samples) and a testing set (10 samples). The class labels of the testing set are assumed unknown first. The logistic classification model will be fitted and validated with the training set and the $K$-fold cross validation method. In [17], Breiman and
Spector mentioned that the 10-fold and 5-fold cross validation work better than leave-one-out for choosing subset of inputs in regression. Also, if $K$ gets too small, the error estimate is pessimistically biased because of the difference in training set size and the full sample size. So, in this case study, the $K=10$ is used for cross validation.

The model fitted and validated is then applied to classify the 10 samples in the testing set. The estimated probabilities of belonging to class A or class B for the testing set are obtained and plotted in Figure 5 (b)-(f). In reality, the first five samples in the testing set are from set A and the other five are from set B. Figure 5 illustrates that logistic regression in combination with Sub, PCA and PLS achieves perfect classification for those 10 testing samples, while logistic regression with Ridge and Lasso misclassified one sample. If the surface samples that have been misclassified with Ridge and Lasso are examined more closely, this surface is actually an outlier because two out of forty local flatness parameters are far larger than the other parameters on that surface. This observation can be explained by the following: (1) for subset selection, the two outlying local flatness parameters are concluded as not significant, so that they are excluded from the classification model. (2) for PCA and PLS, the classification results are not easily influenced by the original flatness parameters because their models are based on the linear combination of the original parameters. Thus, PCA and PLS are not that sensitive to the outliers. (3) For Ridge and Lasso, they shrink all the regression coefficients by penalizing their numbers. If any outlier exists on any surface feature parameters, the classification results will be influenced. So, Ridge and Lasso are more sensitive to outliers than the other methods.
For a more quantitative comparison of the classification accuracies achieved by the five dimension reduction and shrinkage methods, the above analysis are repeated 20 times (randomly splitting the 30 samples into a training set and a testing set for 20 times). Every time, the 10-fold cross-validation is used to validate the model before it is used to classify the testing surfaces. Table 1 shows the means and standard errors of all the 10-fold cross-validated training errors and testing errors for the 20-time repetition using each of the dimension reduction and coefficient shrinkage method. Ridge and Lasso still perform similarly in terms of classification accuracy, while Sub and PLS achieve perfect classification accuracies even with a 20-time repetition. PCA misclassifies 1% of the testing samples. Compared with PLS, PCA is less accurate because it does not take the output variable \( y \) into account when constructing the principle components.
Table 1 Summary of the Misclassification Errors (20 times splitting)

<table>
<thead>
<tr>
<th>Method</th>
<th>10-fold CV Training Error</th>
<th>Testing Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
</tr>
<tr>
<td>OLS</td>
<td>0.160</td>
<td>0.050</td>
</tr>
<tr>
<td>Sub</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PCA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PLS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ridge</td>
<td>0.054</td>
<td>0.029</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.050</td>
<td>0.022</td>
</tr>
</tbody>
</table>

It is noticed from Figure 5 and Table 1 that the classification results from an OLS model for these two sets of pump body surfaces are also provided in this paper. The main purpose is to illustrate the necessity to use the dimension reduction and coefficient shrinkage methods in case the OLS and a couple of surface features are insufficient to classify large and complicated engineering surfaces, while a large set of surface parameters and complicated classification methods are needed. In this case, two summarized overall parameters are used to characterize the surface samples with the OLS. One is the overall flatness $F$, which is local flatness parameters calculated over the entire surface instead of over a small sub-region of the surface. Another parameter is the standard deviation $STD$, defined as the square root of the bias-corrected variance of the height values on the entire surface. The classification decision boundary $p = 1/2$ fitted from the OLS model as well as a scatter plot of all surface samples represented by pairs of inputs $F$ and $STD$ are shown in Figure 6. It seems the linear model is appropriate in this case, since most of the surface samples in set A and set B lie on the different side of the linear decision boundary. However, these samples are also very close to the boundary and sometimes across the boundary. That is why the testing error
using the OLS is high around 23.5%. One way to improve the classification accuracy is to use the non-linear classification models. Another straightforward way is to use more surface feature parameters for characterization, which is the case in this paper. From the results in Table 1, it can be noticed that in terms of classification accuracy the ones using the 40-feature characterization with dimension reduction and coefficient shrinkage methods are generally lower than the one using the 2-feature characterization methods.

It can be concluded that more surface features characterization can capture more surface details for classification purposes in these large and complex engineering surface applications.

Figure 6 Classification results with ordinary logistic regression. Solid line corresponds to \( p = 1/2 \).
Finally, the interpretability of the classification results for process faults diagnosis purposes using the five dimension reduction and shrinkage methods is also analyzed and compared. As mentioned in Section 1, the PCA and PLS are based on the linear combination of the original inputs. Thus, the classification results from these two methods are not interpretable. On the other side, the coefficients of the original inputs are shrunk/selected via the Sub, Ridge and Lasso. The selections of the most significant surface parameters influencing the classification results are achieved with these methods. Four out of forty local flatness parameters (#27, #28, #32, #33) are selected and retained in the logistic model with stepwise regression using the Sub. On the other side, the coefficients estimated using Ridge and Lasso are plotted in Figure 7 (a) and (b). It can be noticed that half of the surface parameters’ coefficients are shrunk to exactly zero using Lasso, which indicates that Lasso selects half of the surface parameters to be included in the classification model. While Ridge doesn’t shrink any coefficients to exactly zeros, the pattern of the coefficient significance is similar to the results from the Lasso. If the significant surface parameters selected from Lasso with non-zero coefficients in Figure 7(b) are mapped onto the original workpiece surface, Figure 7(c) illustrates that the two areas of the pump body surfaces play significant roles in classifying the surfaces machined from A and B, which may help the engineers to identify where, e.g. the fixture setup inside each of the machines, needs to be adjusted in order to obtain consistent surface profile in the future. Comparing these two areas with the parameters obtained via subset selection, it is noticed that subset selection is
computationally simple but slightly aggressive in parameter selection. It only selects one of the areas chosen by the Ridge and the Lasso into its classification model.

Figure 7 Estimated Coefficients from the Training Samples
(a) Ridge (b) Lasso
(c) Significant parameters selected by Lasso
5 Conclusions and Future Work

This paper has researched both the accuracy and interpretability problems of using the dimension reduction and coefficient shrinkage methods in combination with the logistic model to deal with the dimensionality problem in engineering surface classification. Five methods for dimension reduction and coefficient shrinkage are selected and compared in this study. Based on the results of the case study, the two pure dimension reduction methods, the PCA and the PLS, could achieve higher classification accuracies but their results are not interpretable (PCA: 1% error and PLS: 0% error). The Sub could achieve higher accuracy in this case (0% error), but the discrete parameter selection process is aggressive. In comparison, the coefficient shrinkage methods, the Ridge and Lasso, their classification results are interpretable for process faults diagnosis purposes. However, the classification accuracies are generally lower than the other methods (Ridge: 5.8% error and Lasso: 7.3% error).

Since the ability to automatically select the most significant parameters from a large set of the original measurement data and construct an interpretable classification model is considered more important in the modern engineering applications, the selected surface parameters using the Lasso are considered useful even with a 7.3% classification error. The basic objective of the surface classification in this study is to seek the root causes why two sets of pump body surfaces finished from two machines have different local profiles. If one can conclude that 92.7% of the population, which are corrected classified, indicates two major areas on these pump body surfaces are the most significant places responsible for the inconsistency of surface quality, it gives the
engineers enough clue where to go, e.g. to check the fixture setup of each machine, in order to manufacture more consistent surfaces in the future.

Finally, the classification results using the 2-feature characterization and the OLS method are also provided in this study. The conclusion by comparing these results with the ones from using the 40-feature characterization and the dimension reduction and coefficient methods is that using more local features to characterize the surface is a simple and effective way to capture more surface details for classification in large and complex engineering surface applications.

The work in this paper can be extended in some other ways. First, researcher could continue to study how different surface dividing schemes will influence the classification results. In the paper, 40 markers are used to divide each surface into sub-areas, which is just for convenience purpose. Obviously, any numbers for \( p \) whenever \( p > 30 \) can be used. The larger the \( p \) value, the more detailed surface features are supposed to be captured. However, the larger the \( p \) value, the more inaccurate the classification results will be. In addition, the divided sub-areas are also not necessarily to be disjointed. They could be overlapping areas. Second, this study can be applied to a wide variety of engineering surfaces by selecting suitable surface features for each specific application. For example, local surface parameters can be defined from not only the total roughness parameters, but also the average roughness \( R_a \), root mean square roughness \( R_q \), skewness \( R_{sk} \), or kurtosis \( R_{ku} \) for each of the sub-areas. Furthermore, using the logistic model and coefficient shrinkage methods even allow several surface parameters being stacked together to characterize the surface finish for one sub-area in the future study.
Acknowledgement

This work was supported by the Engineering Research Center for Reconfigurable Manufacturing Systems of the National Science Foundation under Award Number EEC-9529125.
REFERENCES

Authors' Biographies

Min Zhang successfully defended her PhD dissertation in the Department of Mechanical Engineering at the University of Michigan – Ann Arbor in Dec. 2006. Her research interests are process faults diagnosis in manufacturing, measurement scheme reduction and engineering surface classification.

Ji Zhu is an assistant professor in the Department of Statistics at the University of Michigan – Ann Arbor. He received his PhD in Statistics from Stanford University in 2003. His research interests are machine learning and data mining, bioinformatics, statistics in finance and marketing, and statistical modeling in engineering and physics.

Dragan Djurdjanovic is an assistant research scientist in the Department of Mechanical Engineering at the University of Michigan – Ann Arbor. He received his PhD in Mechanical Engineering from the University of Michigan in 2002. His research interests are quality control, intelligent proactive maintenance techniques and applications of advanced signal processing in biomedical engineering.

Jun Ni is a Shien-Ming (Sam) Wu Collegiate Professor of Manufacturing in the Department of Mechanical Engineering at the University of Michigan – Ann Arbor. He serves as the director of S.M. Wu Manufacturing Research Center and the co-director of the Multi-Campus National Science Foundation Center for Intelligent Maintenance Systems. His research interests are manufacturing process modeling, analysis and prediction, precision engineering and metrology, cutting tool development, quality control methods, intelligent maintenance systems, monitoring and fault diagnosis.