Efficacy of Some Primary Discriminant Functions in Diagnosing Planetary Gearboxes

Anna Bartkowiak, Radoslaw Zimroz

Efficacy of some primary discriminant functions in diagnosing planetary gearboxes

Anna Bartkowiak[1] and Radoslaw Zimroz[2]

[1] Wroclaw University, Inst. of Computer Science, 50-383 Wroclaw PL
and Wroclaw School of Applied Informatics, 54-239 Wroclaw PL
[2] Wroclaw University of Technology, Diagnostics and Vibro-Acoustics
Science Laboratory, 50-051 Wroclaw PL
and KGHM CUPRUM R&D Center, Mechanical and Electrical
Engineering Group, 53-659 Wroclaw PL

Abstract. We consider the efficiency of some primary discriminant functions applied in planetary gearbox diagnostics. Real data for planetary gearboxes mounted in bucket wheel excavators working in field condition are elaborated. The aim is to perform condition monitoring (faulty or healthy) of such devices. The raw recorded data (vibration series emitted by the device) were first segmented and transformed to frequency domain using power spectra densities (PSD). Next, 15 variables denoting amplitudes of derived spectra were extracted. This yielded two data matrices A and B of size 1232×15, and 951×15, representing the faulty and the healthy device appropriately. The data are non-Gaussian and the covariances in both groups differ significantly.

Now, using Fisher's discriminant criterion and the kernel methodology, we construct from a learning sample (counting only 600 items) a discriminant function able to provide a discriminant score for distinguishing between the healthy and faulty state of a gearbox. The function proved to be very effective: Both for the learning and the testing data samples (600 and 1483 data vectors respectively) we got 100% correct assignments to the 'faulty' and 'healthy' class, with a conspicuous margin between the two classes. The results are visualized in a 2D plane.

Keywords: condition monitoring, discriminant analysis, healthy state, faulty state, kernels, nonlinearity, gearbox diagnostics

1 Introduction

Condition monitoring of the state of a gearbox may be done in a multitude of ways, see the survey papers by [11] or [19] for further references. Gearboxes appear worldwide in many devices and machines. The state of a gearbox is important both for economy and safety reasons. A particular attention is paid to planetary gearboxes which represent a complicated architecture of compounds and are costly in exploitation.

The monitoring of the state of such devices is usually carried out by an analysis of vibration recorded by set of sensors (accelerometers) and specialized data

acquisition systems (Bruel and Kjaer Pulse system was used here). Additionally, auxiliary channel equipped in tachometer (optical probe used often as rotating shaft speed indicator) was used to perform signal segmentation.

Generally the analysis is performed in two phases: *Firstly* the vibration signals are preprocessed and set of diagnostic features is extracted. The aim of this phase is to obtain from the recorded vibration data a kind of (multivariate) numerical data data suitable for further analysis. Features extraction in frequency domain (Fourier analysis like the PSD method) is the basic approach for rotating machinery at this stage. Another method often uses for this purpose wavelet transform, however this is done rather for localized damage than for the considered here distributed form of change of condition. *The second phase* uses various pattern recognition methods, like descriptive statistical methods, neural networks, discriminant functions and classification assignments. Reduction of dimensionality of the data is an important issue in both phases.

A very important point in the analysis is to have proper data for the analysis. For heavy-duty and high power planetary gearboxes the data are difficult to acquire. Therefore quite a lot of researchers use for their analysis data acquired in laboratory conditions using simulations. The behavior of devices working in field conditions might be different from that observed in a laboratory.

Our analysis is based on real data recorded in field conditions. The main contribution is that using these data and applying Kernel Discriminant function (KDF) it is possible to designate a decision boundary allowing for a 100% separation between the faulty and healthy data. The scores of the KDF are visualized in a 2-D plane yielding a visual proof that the two sets of data are really separable.

In the following, Section 2 introduces 3 primary discriminant functions, among them the Fisher's KDF (Kernel Discriminant Function). Section 3 contains a short description of some essential features of the data serving as the basis for our analysis, also some details on constructing the learning and testing samples. In Section 4 results of our analysis, when applying the 3 discriminant function, are presented. Section 5 contains some discussion and closing remarks.

2   Some primary discriminant functions

There is a multitude of discriminant functions based on diversified criterions. We consider only the two group case when the data are subdivided into two classes (like: 'faulty' or 'healthy') represented by two groups of the observed data. The sought discriminant function should produce for each data vector a score (called also decision score) allowing to assign that vector to one of the considered classes. We will consider only some *basic* discriminant functions which - to our opinion - are advised to start with, when attempting to build a discriminant function providing a decision boundary between two groups of observed data. In next subsections we consider in turn:

(i) Linear discriminant function formulated and solved as ordinary multivariate regression function with predicted values $y = +1$ for one of the groups and $y = -1$ for the other group.
(ii) Classical canonical discriminant function based on Fisher's criterion (called often FDA, Fisher's discriminant analysis).
(iii) Discriminant functions using kernels providing a nonlinear approach.

Other methods of discriminant analysis - and there are many of them - may be found, among others, in [9, 10, 16].

Denotations:

$\diamond$ $\mathbf{X} = \{x_{ij}\}$, $i = 1, \ldots, n$, $j = 1, 2, \ldots, d$ — the recorded matrix of observations, with d columns denoting variables, and n, the number of rows of the matrix. Each row of $\mathbf{X}$ constitutes a data vector $\mathbf{x}_i$ containing data recorded for the ith individual, called also data item. The vector $\mathbf{x}_i$ is written as the column vector $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T$; it may be also considered as point located in the data space R^d. Each of the data vectors $\{\mathbf{x}_i\}$ may belong to only one of the two groups of data.

$\diamond$ $\mathbf{y} = \{y_i\}$, $i = 1, \ldots, n$ – the vector of group labels. Each data vector has its label indicating to which group (class) the given (ith) data vector is belonging. The group membership is crisp: each data vector may belong only to one of the two considered classes.

2.1   Linear discriminant function via linear regression analysis

We consider here the regression model

$$y_i = b_1 x_{i1} + b_2 x_{i2} + \ldots + b_d x_{id} + b_0 + e_i, \quad i = 1, \ldots, n \quad (1)$$

The vector $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T$ denotes here the predicted variable, which takes specifically only two values: -1, if the ith data vector comes from the faulty class, and $+1$, if the ith data vector comes from the healthy class.

The coefficients $b_1, \ldots, b_d, b_0$ appearing in (1) are called regression coefficients, they are usually unknown and have to be estimated in a supervised way, that is from a learning sample, which contains known values od the data matrix $\mathbf{X}$ and of the vector $\mathbf{y}$. The regression coefficients are commonly estimated using the principle of Least Squares (LS), which minimizes the following criterion:

$$SS_e = \sum_{i=1}^n (e_i)^2 = \sum_{i=1}^n (y_i - b_1 x_{i1} - \ldots - b_d x_{id} - b_0)^2 .$$

To obtain the solution, that is, the vector $\mathbf{b} = [b_1, b_2, \ldots, b_d, b_0]^T$ which yields the minimum of the quadratic form SS_e, we have firstly to define $\mathbf{X}^+$, the augmented data matrix $\mathbf{X}$, as:

$$\mathbf{X}^+ = [\mathbf{X}, \mathbf{1_n}],$$

where $\mathbf{1_n}$ denotes a column of n ones.

With known $\mathbf{X}^+$ and $\mathbf{y}$, the estimates of $\mathbf{b}$, denoted in the following as $\hat{\mathbf{b}}$, are obtained as the solution of the following set of linear equations:

$$(\mathbf{X}^+)\,\hat{\mathbf{b}} = \mathbf{y}, \tag{2}$$

In Matlab the solution $\hat{\mathbf{y}}$ may be obtained simply via the matlab left divide function.

To obtain $\hat{\mathbf{y}}$, the predicted scores of class assignment of vectors contained in a data matrix say $\mathbf{X}_{learn}$ we add to that matrix a column of ones $\mathbf{1}_n$ obtaining the augmented data matrix $\mathbf{X}_{learn}^+$, and perform the multiplication

$$\hat{\mathbf{y}}_{learn} = (\mathbf{X}_{learn}^+)\,\hat{\mathbf{b}}.$$

Similarly, using the vector $\hat{\mathbf{b}}$ yielded by the learning sample, we may obtain the predicted values $\hat{\mathbf{y}}_{test}$ for the test sample $\mathbf{X}_{test}^+$ as

$$\hat{\mathbf{y}}_{test} = (\mathbf{X}_{test}^+)\,\hat{\mathbf{b}}.$$

Predicted values $\hat{\mathbf{y}}$ near to -1 (generally: negative values) indicate similarity to the class 'faulty', and values of $\hat{\mathbf{y}}$ near to $+1$ (generally: positive values) indicate similarity to the class 'healthy'.

2.2   FDA, classical canonical discriminant analysis based on Fisher's criterion

The BW-projections using the between and within groups scatter. The main idea of the method is to find a vector $\mathbf{u} = (u_1, ..., u_d)^T$ serving for construction of a new variable (feature) Z with values z_i obtained from the observed data vector $\mathbf{x}_i$ in the following manner:

$$z_i = \mathbf{x}_i^T \mathbf{u}. \quad i = 1, \ldots, n. \tag{3}$$

The values of the new variable, denoted as z_i, $i = 1, \ldots, n$, may be considered as projections of the data vectors $\mathbf{x}_i$ onto a new axis called Z.

In the following we will consider the case when the the entire set of data vectors is subdivided into $G = 2$ groups (classes). For the ith data point of the gth group ($i = 1, \ldots, n_g$, $g = 1, 2$), we denote the projection of the data point $\mathbf{x}_i^{(g)}$ as $z_i^{(g)}$, and the general transformation may be read as:

$$z_i^{(g)} = (\mathbf{x}_i^{(g)})^T \mathbf{u}.$$

Thus, we obtain G=2 groups of projections:

$$\underbrace{z_1^{(1)}, ..., z_{n_1}^{(1)}}_{\text{1st group}}, \underbrace{z_1^{(2)}, ..., z_{n_2}^{(2)}}_{\text{2nd group}} \tag{4}$$

For values of the variable Z appearing in formula (4) above we may calculate group means $\bar{z}.^{(g)}$, ($g = 1, 2$), and the overall mean $\bar{z}.^{(\cdot)} \equiv \bar{\bar{z}}$ (this is the mean of

all data vectors, without subdivision into the 2 groups). Moreover, basing on the values given in (4), we may define and calculate ZW, the *within group scatter* of the variable Z, and ZB, the *between group scatter* of the variable Z, as

$$ZW = \sum_{g=1}^{2}\sum_{i=1}^{ng}(z_i^{(g)} - \bar{z}.^{(g)})^2, \quad ZB = \sum_{g=1}^{2}n_g(\bar{z}.^{(g)} - \bar{\bar{z}})^2.$$

Which projections are the best for the purpose of discriminant analysis? Consider the following issues. For a good discriminant function
⋄ points-projections *belonging to different groups* of data *should be separated as much as possible*,
⋄ points-projections *belonging to the same groups* should be possibly much concentrated around their group-means.

Taking this into account, R.A. Fisher has formulated the following criterion, which is called today **Fisher's criterion**:

$$J(z) = \frac{ZB}{ZW} = \frac{\sum_{g=1}^{2}n_g(\bar{z}.^{(g)} - \bar{z}.^{(\cdot)})^2}{\sum_{g=1}^{2}\sum_{i=1}^{n_g}(z_i^{(g)} - \bar{z}.^{(g)})^2}. \tag{5}$$

To indicate for some differentiation between the groups, the criterion $J(z)$, evaluated for real data, should have a value larger than 1; the larger the better.

To find the best vector $\mathbf{u}$ maximizing the criterion $J(z)$ we express (5) in terms of $\mathbf{u}$ and $\mathbf{X}$; usually $\mathbf{X}$ is centered to zero, that is, all column means are equal zero. After some algebra manipulations we obtain

$$J(z) = J(\mathbf{u}; \mathbf{X}) = \frac{\mathbf{u}^T\mathbf{Bu}}{\mathbf{u}^T\mathbf{Wu}}, \tag{6}$$

where

$$\mathbf{B}_{d \times d} = \sum_{g=1}^{2}n_g(\bar{\mathbf{x}}.^{(g)} - \bar{\mathbf{x}}.^{(\cdot)})(\bar{\mathbf{x}}.^{(g)} - \bar{\mathbf{x}}.^{(\cdot)})^T, \tag{7}$$

$$\mathbf{W}_{d \times d} = \sum_{g=1}^{2}\sum_{i=1}^{ng}(\mathbf{x}_l i^{(g)} - \bar{\mathbf{x}}.^{(g)})(\mathbf{x}_i^{(g)} - \bar{\mathbf{x}}.^{(g)})^T. \tag{8}$$

The above equations show clearly that the criterion J may be considered as a function of the sought unknown vector $\mathbf{u}$ and the known data matrix $\mathbf{X}$, which serves for calculations of the between and within group scatter matrices $\mathbf{B}$ and $\mathbf{W}$, each of size $d \times d$. Applying Lagrange's methodology and using the additional restraint $\mathbf{u}^T\mathbf{Wu} = 1$, one arrives to the two-matrices eigenvalue problem:

$$(\mathbf{B} - \lambda\mathbf{W})\mathbf{u} = \mathbf{0}. \tag{9}$$

which is well known in matrix algebra. It is shown, that for $G = 2$ groups there is an unique vector $\mathbf{u}$ satisfying eq. (9). Moreover, the constant λ is equal to the quotient $J(z)$ appearing in formula (5), evaluated at the solution $\mathbf{u}$. In other

words, the constant λ is equal to the ratio of the between group scatter divided by the within group scatter defined above. It satisfies the inequality:

$$\lambda \geq 0$$

Large values of λ ($\gg 1.0$) indicate a big differentiation between projections of the two groups of data.

Using the obtained vector $\mathbf{u}$ for the projection $z = \mathbf{x}^T \mathbf{u}$ we obtain a new variable Z satisfying Fisher's criterion (5). The technique of constructing such variables is called *Canonical Discriminant Analysis* (*CDA*), the obtained projections Z are called *Canonical variates* or *CDA scores*. The CDA scores may be calculated and displayed both for the learning and testing samples of data.

The BT projections using the between groups and total scatter.

The criterion $K(z)$ formulated defined in formula (5) might be formulated in another way. Let T denote the total data scatter matrix:

$$\mathbf{T}_{d \times d} = \sum_{g=1}^{2} \sum_{i=1}^{ng} (\mathbf{x}_l i^{(g)} - \bar{\mathbf{x}}.^{(\cdot)})(\mathbf{x}_i^{(g)} - \bar{\mathbf{x}}.^{(\cdot)})^T, \tag{10}$$

with $\bar{\mathbf{x}}.^{(\cdot)}$ denoting the overall mean calculated from all the data.

Now our aim is to find the vector $\mathbf{v}$ maximizing the following index denoting the ratio of the between group to the total scatter

$$\eta = \frac{\mathbf{v}'\mathbf{B}\mathbf{v}}{\mathbf{v}'\mathbf{T}\mathbf{v}} \tag{11}$$

The solution of the problem (11) is obtained in a similar way as that of (5); it leads to the two-matrix problem

$$(\mathbf{B} - \eta\mathbf{T})\mathbf{v} = \mathbf{0}. \tag{12}$$

Now the value *eta* evaluated at the solution $\mathbf{v}$ of (12) takes values from the interval

$$0 \leq \eta \leq 1.0$$

Big values of η indicate for a good separability of the groups.

Similarly as for Fisher's projections using the $\mathbf{B}$ and $\mathbf{W}$ matrices, the values z_i of the *BT-projections* are obtained as:

$$z_i = \mathbf{x}_i^T \mathbf{v}, \tag{13}$$

where $\mathbf{v}$ is the solution of (12). The projections z_i obtained that way may be visualized similarly as those obtained from the BW-projections.

Let us emphasize: The canonical discriminant variables represent projections to lower dimension subspaces. They serve in first place for visualization of the projected points. The differentiation of the projected points follows from the assumed Fisher's criterion. The group assignment is relatively easy only in the 2-group case. Having $G > 2$ groups, one should use for class assignments a specialistic tool, e.g. Mahalanobis distances or k-nearest neighbor methods.

2.3 Discriminant functions using kernels

Kernel methods may be defined and used in a number of ways, see for example [15, 13, 14, 6–8]. One might say shortly that this is canonical discriminant analysis carried out in an extended space $\mathcal{F}$ obtained by a non-linear mapping of the original data. The applied mapping takes into account various nonlinear relations between the observed variables which makes that in the extended feature space $\mathcal{F}$ the classical algorithms (like CDA) become more powerful.

Let φ denote the transformation carrying out the mapping of the observed vector $\mathbf{x}$ ($\mathbf{x} \in R^d$) to the extended feature space $\mathcal{F}$. The function φ is called the mapping function. The extended space $\mathcal{F}$ must have defined an operator called scalar product for every pair of its elements $\varphi(\mathbf{x}_i)$, and $\varphi(\mathbf{x}_j)$. The scalar product is defined via so called dot product for pairs of data points obtained as images of the mapping φ of points $\mathbf{x}_i$ and $\mathbf{x}_j$ from R^d to $\varphi(\mathbf{x}_i)$, and $\varphi(\mathbf{x}_j)$ in $\mathcal{F}$:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)), \tag{14}$$

where the symbol '·' means the scalar product between the vectors-images $\varphi(\mathbf{x}_i)$ and $\varphi(\mathbf{x}_j)$ evaluated in $\mathcal{F}$. Let us point out, that the kernel matrix $\mathbf{K}$ is of size $n \times n$, with n being the size of the learning sample, thus its evaluation may be computationally demanding.

Performing the mapping $\varphi(\mathbf{x})$ might be quite cumbersome – this depends from the applied mapping φ, in particular from the dimension of $\mathcal{F}$ which is usually much larger than that of R^d. However, by looking at the transformations carried out using particular mapping functions φ it was found that the elements of the kernel matrix $\mathbf{K}$ defined in (14) may be evaluated directly from the points $\mathbf{x}_i$ and $\mathbf{x}_j$ located in R^d *without* considering their images in $\mathcal{F}$. Then, if we are able to formulate the solution of our task in terms using only the elements of the kernel matrix $\mathbf{K}$, we are the winners: we do not need anymore the mapped points from $\mathcal{F}$. This possibility is called 'the kernel trick'.

For example, this is true when considering as φ the Gaussian or polynomial kernels and having as task canonical discriminant analysis or principal component analysis. Detailed algorithms how to do it in the case of discriminant analysis for G groups of data are shown, e.g., in [14] (when using the scatter matrices $\mathbf{B}, \mathbf{W}$) or [6] (when using the scatter matrices $\mathbf{B}, \mathbf{T}$). The algorithms are somehow lengthy and we do not show them here.

More details on the methods and examples of applications may be found in [6–8, 2–4].

3 Experiments with the gearbox data

In the following we will show an analysis conducted using true data, that is recorded from machines working in field conditions. The data were recorded by Bartelmus and Zimroz [1]. We had 2 groups of data contained in 2 data matrices A of size 1232×15 and B of size 951×15 obtained from vibration data of two

gearboxes, the one in a faulty and the other in a healthy condition. The data were the basis of an elaboration presented in [1].

The raw recorded data (vibration series of the devices) were firstly segmented and transformed to frequency domain using power spectra densities. Next, 15 variables denoting amplitudes of derived spectra were obtained. This yielded two data matrices (A and B of size 1232×15, and 951×15 appropriately) characterizing the state of the faulty and the healthy device. The rows of the matrices constitute data vectors (called also data items) living in the 15-D space R^{15}. Thus we have two samples of faulty and healthy data containing n1=1232 and n2=951 data items appropriately.

Taking as a new feature the sum of all the 15 variables, Bartelmus and Zimroz [1] were able to classify – on the base of the proposed feature – about 80 % of all data vectors. To classify the remainder, they needed an external variable indicating for the actual load of the working excavator.

The data were more thoroughly investigated in [17, 5, 18]. It appeared that the distribution of the variables is not Gaussian, the data contains a considerable number of outliers, moreover, the covariance structure in the two groups (faulty and healthy) is markedly different. It became evident, that a proper analysis of the data needs non-linear methods [18].

Now we have subdivided the entire data set into the learning and testing samples. The learning sample was obtained from randomly chosen 300 rows of the A matrix and other 300 rows of the B data matrix; this yielded the learning sample of size 600×15. The remainder of the data (932 rows from A and 651 rows from B) yielded the testing sample of size 1583×15.

4   Results

4.1   Results when using linear regression method

The linear regression was supposed to predict the class score given by the values:
$y_i = -1$ for data vectors i belonging to the faulty class, and
$y_i = +1$ for data vectors i belonging to the healthy class.

The calculations were done in Matlab. The estimates of the regression coefficients $\hat{\mathbf{b}} = [\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{15}, \hat{b}_0]^T$ were obtained using the learning sample and the Matlab function mldivide, as described in Section 2.1.

The obtained estimates $\hat{\mathbf{b}}$ were used to obtain the predicted values of the regression. This yielded the predicted vectors of discriminant scores $\hat{\mathbf{y}}_{learn}$ for the learning and $\hat{\mathbf{y}}_{test}$ for the testing samples. The obtained scores are visualized in Figure 1.

We have used the following decision rule:
If for a data vector $\mathbf{x}_i$ its predicted value satisfies $\hat{y}_i \leq 0$, classify that data vector as 'faulty',   otherwise — classify it as 'healthy'.

Looking at the results shown in Fig. 1 we find that:
In the *learning sample* we have 3 incorrect assignments for a total of 600 items. The erroneous assignments happened in the faulty group, where 3 faulty items
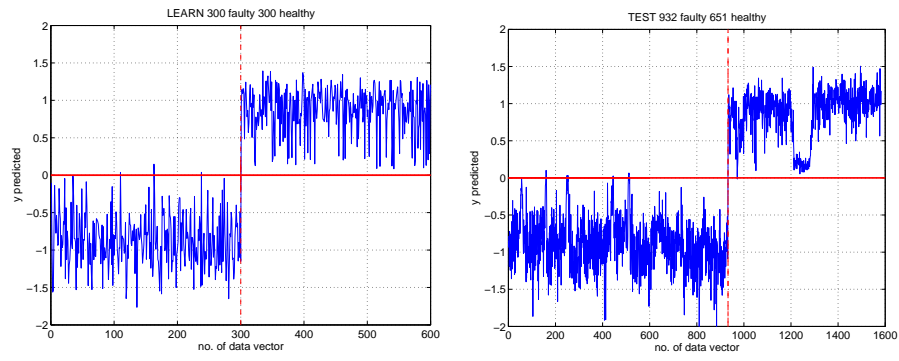
Fig. 1. Discriminant scores obtained via linear regression method with target values $y = -1$ for the faulty and $y = +1$ for the healthy state of the machine. Left panel: learning sample, first 300 faulty and next 300 healthy items. Right panel: testing sample, first 932 faulty and next 651 healthy items.

have been assigned to the healthy group. This makes the **error rate equal to 0.0050**.

In the *testing sample* we have 11 incorrect assignments for a total of 1583 items: 10 faulty items were assigned as healthy, and 1 healthy item was assigned as faulty. This makes the **error rate equal to 0.0069**.

Thus the linear discriminant function is a good choice for using it as a discriminant function, especially as it is very easy to carry out.
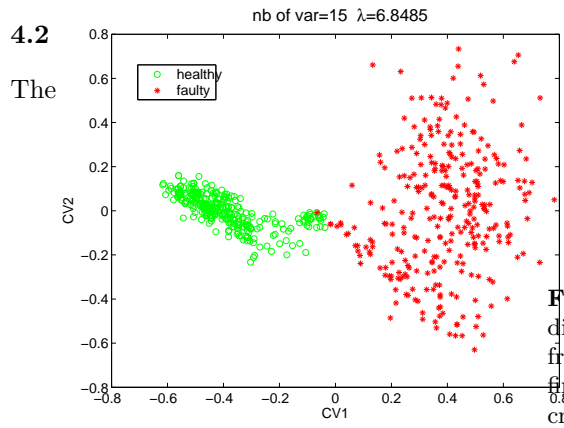
4.2 analysis

The jections - is shown in Figure 2.

Fig. 2. Display of two canonical discriminant functions constructed from d=15 dimensional data. The first canonical variate CV1 has discriminative power $\lambda = 6.8485$. The second CV2 has discriminative power equal to 0, however is helpful in perceiving the spread of the data points in the two groups.

We have calculated canonical discriminant functions CV1 and CV2 using the BW-projections based on Fisher's Criterion (5) and using own software programmed in Matlab. The main result is shown in Figure 2. Looking at that figure one can grasp at once how different is the spread of points in the two groups of data.

The first canonical variate CV1 has discriminative power lambda=6.8485. We have constructed a second canonical variate which has discriminative power equal to 0, however helps in visualization of the two data clouds representing the faulty and healthy data points.

Concerning class assignments, we have the following results:
Learning sample: 3 erroneous assignments in the faulty class. Thus the error rate equals $3/600 = 0.005$.
Testing sample: There are 24 erroneous assignments in the faulty class. Thus the error rate equals $24/1583 = 0.0154$.

4.3   Results of Kernel canonical discriminant analysis

The calculations were performed using BT projections elaborated by Baudat and Anouar [6] and using their software gda downloaded from www.kernel-machines.org. As previously, the entire observed data were subdivided into the learning sample and test sample – the same, as used in subsections 4.1 and 4.2. The original d=15 dimensional data were mapped to the feature space $\mathcal{F}$ where the kernel matrix $K = \{k(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1, \ldots, n\}$ was constructed from the data vectors belonging to R^d using the Gaussian kernels with parameter $\sigma = 0.7$:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j))/\sigma \tag{15}$$

In such a way a non-linear mapping of the original data into $\mathcal{F}$ was obtained.

The proper analysis was carried out in the extended space $\mathcal{F}$. The vector $\mathbf{v}$ performing the BT-projections onto new axis Z positioned now in $\mathcal{F}$ was obtained using the kernel matrix $\mathbf{K}$ whose elements were calculated using formula (15). Next, using only information from the kernel matrix $\mathbf{K}$, the projection vector $\mathbf{v}$ was constructed. This allowed us to construct new variable Z satisfying the criterion (11), however valid for the mapped data in $\mathcal{F}$. Next, using the functions spreadGda and plotGda (see [6]), it was possible to recover the projections $z_i, i = 1, \ldots, n$ for each individual data vector from the learning sample ($n_{learn=600}$) and test sample ($n_{test} = 1283$). The obtained scores z_i obtained now are analogous to the scores z_i defined for the classical canonical discriminant analysis (see eq. (13)) however the obtained now (from spreadgda and plotgda) are projections from the non-linear mapping valid in $\mathcal{F}$. They are shown in Figure 3. One may notice that **the projections** from the faulty and healthy data vectors **are completely disjoint**. The criterion η amounts $\eta = 0.9938$.
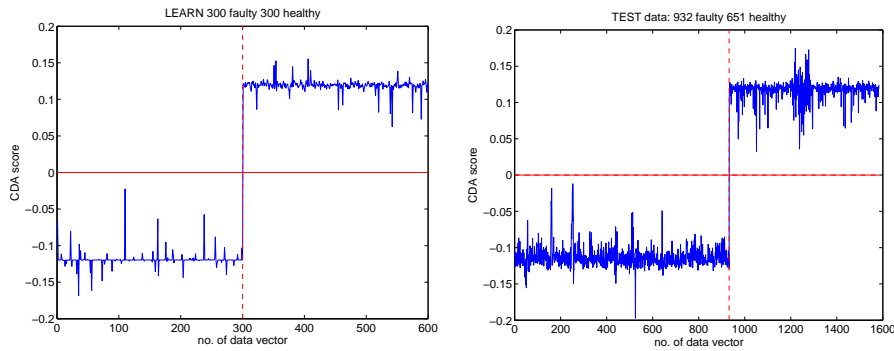
Fig. 3. Kernel discriminant function obtained from Gaussian kernel with $\sigma = 0.7$. Negative and positive scores indicate the faulty and healthy state of the machine. Left panel: learning sample, first 300 faulty and next 300 healthy items. Right panel: testing sample, first 932 faulty and next 651 healthy items. Subsequent data items are linked by line segments. All items in both of the classes are classified perfectly.

5 Discussion and Concluding Remarks

Using real data gathered in field conditions we have investigated the usefulness of some primary discriminant functions in diagnosis of the faulty or healthy state of two planetary gearboxes. Thanks to applying a nonlinear transformation of the gathered data (using Gaussian radial basis functions) to an extended feature space $\mathcal{F}$ we have obtained a perfect decision boundary between the faulty and healthy data. The function designating decision boundary was obtained from a relative small learning sample counting 300 faulty and 300 healthy data vectors. The behavior of the obtained decision boundary was checked using a larger test sample counting 1583 data vectors. All of them were correctly classified to the proper group of their origin. We were also able to represent the projections from the extended feature space in an ordinary 2-D plane, as shown in figs. 1 and 3.

The elaborated data were not-normal and with covariance matrices differing substantially [17, 18]. Despite that the constructed discriminants have worked surprisingly effectively.

References

1. Bartelmus, W., Zimroz, R.: A new feature for monitoring the condition of gearboxes in nonstationary operating systems, Mechanical Systems and Signal Processing 23 (5), 1528-1534 (2009).
2. Bartkowiak, A., Evelpidou, N.: Visualizing Some Multi-Class Erosion Data Using Kernel Methods. A. Rizzi and M. Vichi (Eds), Proceedings in Computational Statistics, 17th Symposium Held in Rome, Contributed paper, pp. 805–812. Physica Verag, A Springer Company (2006).

3. Bartkowiak, A., Evelpidou, N.: Visualizing of some multi-class erosion data using GDA and supervised SOM. K. Saeed et al. (Eds), Biometrics, Computer Security Systems and Artificial Intelligence Applications. Springer, pp. 13–22 (2006).
4. Bartkowiak, A., Evelpidou, N., Vasilopoulos, A.: Visualization of Five Erosion Risk Classes using Kernel Discriminants. In: J. Pejaś, K. Saeed (Eds), Advances in Information Processing and Protection, ISBN: 978-0-387-73136-0 (Print) 978-0-387-73137-7 (Online) Chapter 10, Springer, pp. 169-178 (2007).
5. Bartkowiak, A., Zimroz, R.: Outliers analysis and one class classification approach for planetary gearbox diagnosis. Journal of Physics: Conference Series 305 (1), art. no. 012031 (2011).
6. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. Neural Computation **12**, 2385–2404 (2000).
7. Camastra, F.: Kernel Methods for Computer Vision, Theory and Applications. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.9315
8. Camastra, F. : Kernel Methods for unsupervided Learning. PhD thesis. DISI-TH, Universita di Genova, 2004.
9. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd Edition. Wiley (2001)
10. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning; Data Mining, Inference and Prediction. 2nd Edition, New-York, Springer. (2010)
11. Jardine, A.K.S., Lin, D. Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process. 20, 2006, 1483-1510.
12. Liu, Z., Qu, J., Zuo, M.J., Xu, H.B.: Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and kernel Fisher discriminant analysis. Int. J. Adv. Manuf Technol., DOI 10.1007/s00170-012-4560-y
13. Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Smola, A., Muller, K-R.: Constructing descriptive and discriminative nonlinear features: Rayleigh coefficients in kernel features spaces. IEEE Trans. PAMI 25(5) 623-628 (2001).
14. Műller K-R., Mika S., Rätsch G., Tsuda K., Schőlkopf B.: An introduction to kernel-based learning algorithms. IEEE Trans. on Neural Networks, 12 (2), 181–202 (2001).
15. Shawe-Taylor, J., and Christianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press UK. (2004).
16. Trendafilov, N., Vines, K.: Simple and interpretable discrimination. Comput. Stat. Data Anal. 53 979-989 (2009).
17. Zimroz, R., Bartkowiak, A.: Investigation on spectral structure of gearbox vibration signals by principal component analysis for condition monitoring purposes. Journal of Physics: Conference Series 305 (1), art. no. 012075, (2011).
18. Zimroz, R., Bartkowiak, A.: Multidimensional data analysis for condition monitoring: features selection and data classification. CM2012—MFPT2012. BINDT, 11-14 June, London. Electronic Proceedings, 2012, art no. 402, pp. 1-12 (2012).
19. Zimroz, R., Bartkowiak,A.: Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions. Mech. Syst. Signal Process. 38(1), 237-247 (2013).