Investigation on Spectral Structure of Gearbox Vibration Signals by Principal Component Analysis for Condition Monitoring Purposes

Radoslaw Zimroz¹ and Anna Bartkowiak²
¹ Wroclaw University of Technology, Diagnostics and Vibro-Acoustics Science Laboratory, PL
² University of Wroclaw, Institute of Computer Science, Wroclaw, PL
E-mails: radoslaw.zimroz@pwr.wroc.pl, aba@ii.uni.wroc.pl,

Abstract. Spectral analysis is well-established analysis of vibrations used in diagnostics both in academia and industry. In general, one may identify components related to particular stages in the gearbox and analyze amplitudes of these components with a simple rule for decision-making: if amplitudes are increasing the condition becomes worse. However, usually one should analyze not single amplitude but at least several components, but: how to analyze them simultaneously? We have provided an example (case study) for planetary gearboxes in good and bad conditions (case B and case A). As diagnostic features we have used 15 amplitudes of spectral components related to fundamental planetary mesh frequency and its harmonics. Using Principal Component Analysis (PCA), it has been shown that amplitudes don’t vary in the same way; change of condition affects not only amplitudes of all components in that sense, but also relation between them. We have investigated geometry of the data and it has been shown that the proportions of the explained total inertia of the three data sets (“good”, “bad” and mixed good/bad) are different. We claim that it may be a novel diagnostic approach to employ multidimensional analysis for accounting not only directly observed values but also interrelations both within and between the two groups of data. Different structure of the data is associated with different condition of the machines and such assumption is specified for the first time in the literature. Obviously it requires more studies.

1. Introduction
Condition Monitoring of gearboxes is one of the most exploited problems in the field. In recent years some interesting approaches based on multidimensional data analysis have appeared [1-7,14]. Until now, some researchers noticed that PCA may be used for reduction of data dimensionality [8, 10,14]. The PCA technique has been applied together with other statistical techniques for time domain data (like RMS, kurtosis, skewness, peak etc), time-frequency parameters (for example wavelet coefficient [8,10,11]) and other [9,16]. PCA has been applied recently also in data processing for time varying systems (for structural health monitoring [15] and for planetary gearbox diagnosis [19]). In this paper it is proposed to use features extracted from Power Spectral Density of raw vibration signals captured from planetary gearbox working under time-varying load conditions [12]. Obviously, non-stationarity of the load causes a lot of troubles with application of spectral analysis. Most of examples in the literature propose to use order analysis instead of classical spectral analysis [13], however, spectral smearing has been overcome here by signal segmentation. We have assumed that for short segments
signal is locally stationary, moreover, due to limited resolution smearing is neglectable (see figure 2a, one may easily notice family of equally spaced components related to mesh frequency and harmonics). Two machines have been analyzed: one in good and the other in bad condition. After feature extraction it has been found using PCA that these two data sets have different structure and real dimension of data differs as well. This result provides a new idea that these differences (structure, dimension) can be used as diagnostic features. Our thesis is that the two data sets differ in their internal structure, and that this may be stated by making a spectral decomposition of the correlation matrices of the gathered data. In order to prove it, the paper is organised as follows: in next section we will briefly describe machines, measurements and feature extraction procedure, next theoretical background of PCA will be presented (how to make the spectral decomposition of a square Gramian matrix, and some formal properties of the obtained decomposition - useful for stating the sought differences). Finally, chapter 4 provides results of analysis related to: dimension of data, distribution of total inertia of data, structure of data expressed by first 3 PCs and correlation between features from original data and their potential for diagnostics.

2. Experiment description. Feature extraction.

Two experiments have been conducted. Two multistage gearboxes with planetary stage have been investigated; they are used for driving the bucket wheel in bucket wheel excavators working in lignite surface mining. Figure 1(a) shows general view of one of the investigated machines. Analysed gearboxes were in different conditions (according to observations provided by engineers) and have been used for 20 000h (bad condition) and 10 000h (good condition). After dismantling of the gearbox with bigger lifetime, comments provided by manufacturer were as follows: all rolling elements bearings have over-limit radial backlash so have to be replaced by new ones, almost all gears should be improved by grinding and scuffing on the teeth and micro-cracks have also been spotted [12]. The same measurement procedure has been applied to the two investigated gearboxes. In order to measure diagnostic signals, the Bruel&Kjaer Pulse system has been used. During the experiments data from six channels have been acquired: two auxiliary channels (tachometer signal and electric current signal) and four channels of vibration data. Duration of each signal was 60s and sampling frequency was set to 16384Hz. The tachometer was mounted on the input shaft of the gearbox. Location of accelerometers (figures 1.b,c,d) and scheme of gearbox (figure 1d) are presented in figure 1.

Figure 1. The investigated machine: a) general view b), c) location of accelerometers, d) scheme of machine
For data acquisition, the speed profile and the vibration signal data were processed. The vibration sensor was located as close as possible to the planetary stage and positioned in the vertical direction. The processing (figure 3, top part) included signal segmentation (according to digging process cyclicity) and feature extraction, which yielded the mean speed for each segment obtained from the speed profile, and 15 amplitudes of the planetary mesh frequency components, obtained from the spectral analysis of vibration. In this paper we use only features extracted from one vibration channel (channel S3).

Figure 2 a) The idea of features extraction from Power Spectral Density of planetary gearbox vibration signal b),c). Examples of real Power Spectral Density for planetary gearbox vibrations

Figure 3 A diagram of data preparation and feature processing using PCA and correlation analysis
All acquired data (speed profile and vibration signal) have been submitted to a procedure presented on upper part of figure 3, so finally two matrices of data have been formed: \( A \) and \( B \). These are rectangular matrices of size \( n \times d \) with \( n \) denoting the number of rows and \( d \) the number of columns of the data matrix. For the investigated data \( d=15 \) (\( d \) means the number of spectral components, called in the following also PP1, \ldots, PPd), and the number of rows equals \( n_a=1232 \) and \( n_p=951 \) respectively. The symbols \( A \) and \( B \) denote bad and good condition of the given machine, respectively. These two matrices have been fused together into one matrix \( C \) with \((n_a+n_p)\) rows and \( d = 15 \) columns. Second part (bottom) of figure 3 presents an analysis scheme and it will be discussed later.

3. Spectral decomposition and PCA for a covariance matrix \( S \) – theoretical background

The spectral decomposition of a covariance or correlation matrix is widely used method for multidimensional data processing [17]. In the statistical community the method is called Principal Component analysis (PCA), while the signal processing community refers to the method as the Karhunen–Loève transform. The method is defined as a linear transformation of original multidimensional data to new coordinate system being generally of lower dimension.

Let \( \textbf{X} \) denote the analyzed data matrix, with \( n \) denoting the number of instances, that is data vectors with \( d \) components each (in our case segments of the vibration signal). Then principal components (new variables constructed as linear combinations of the original, observed variables) are defined as \( \textbf{Y} = \textbf{X} \textbf{A} \). The column vectors composing the matrix \( \textbf{A} \) are found from the secular matrix equation:

\[
\left( \textbf{S} - \lambda \textbf{I} \right) \textbf{a}_i = 0,
\]

for \( \lambda \) denoting the covariance (or correlation) matrix of \( \textbf{X} \) (in our case: data matrix \( \textbf{A} \) or \( \textbf{B} \)). In the following we will perform PCA using correlation matrices. For a positive definite covariance matrix \( \textbf{S} \) we obtain as solutions \( d \) vectors \( \textbf{a} \) satisfying the secular matrix equation; the solutions appear in pairs \( (\lambda, \textbf{a}) \) with the eigenvalues \( \{\lambda_i, i=1, \ldots, d\} \) ordered in descending order. The column vectors \( \{\textbf{a}_i\} \) span the principal component space.

The new values composing the columns of the derived matrix \( \textbf{Y} \) of size \((n \times d)\) are called principal components; they constitute new features characterizing the data. In the following we will denote them as PC1, PC2, \ldots, PCd. The new features (that is, the PCs) are uncorrelated.

The derived eigenvalues and eigenvectors constituting the principal components have the following properties:

**Property 1. Theorem on the spectral decomposition of a covariance matrix.** The full rank covariance matrix \( \textbf{S} \) of size \( d \times d \) may be composed from \( d \) unitary matrices of size \( d \times d \), each of rank 1, as: \( \textbf{S} = \lambda_1 \textbf{a}_1 \textbf{a}_1^T + \lambda_2 \textbf{a}_2 \textbf{a}_2^T + \ldots + \lambda_d \textbf{a}_d \textbf{a}_d^T \). Considering traces of the left hand and right hand matrices of the above equality it can be shown that \( s_{11} + s_{22} + \ldots + s_{dd} = \lambda_1 + \ldots + \lambda_d \), with \( s_{jj} \) denoting the variance of the \( j \)-th variable.

The sum of all the variances \( s_j \) (and eigenvalues \( \lambda_j \)) is often called the total inertia of the data.

**Property 2. Variances of the constructed PCs.** The variances of successive PCs are equal to the respective eigenvalues: \( \text{Var}(\text{PC}_j) = \text{Var}(y_j) = \lambda_j \), \( j = 1, \ldots, d \), with \( y_j \) denoting the \( j \)-th column of the matrix \( \textbf{Y} \).

**Property 3. Possibility of data reduction and data visualization.** Obviously, when the first eigenvalues are big and constitute the majority of the total inertia, then the variances of the corresponding PCs constitute the majority of the total inertia; it is said that they reproduce the majority of total inertia of the data. In such a case a few PCs may represent a big number of original variables, which means a reduction of dimensionality of the original data set. In particular, it may happen that the first two or three PCs reproduce the majority of the total inertia. In such a case the original high-dimensional data may be transformed to 2- or 3-dimensional space and effectively visualized in that space.

Next section shows results from applying the PCA to the data sets \( A \), \( B \), and \( C \). In particular, the aspects of dimensionality reduction and data visualization will be elaborated.
4. Results of analysis

Our thesis formulated in the introduction was that the two data sets differ in their internal structure, and that this may be stated by making a spectral decomposition of the correlation matrices of the gathered data. In this chapter we will provide results of analysis related to dimension of data, distribution of total inertia of data, structure of data expressed by first 3 PCs, correlation between features from original data and their potential for diagnostics.

4.1. Analysis of eigenvalues

The three gathered data sets (A, B, and C) have been analysed with the aim to find real, intrinsic dimension of these data and to prove that it is different for gearboxes in good and bad conditions. First of all, eigenvalues (e-vals) derived for the sets A, B, and C, when using correlation matrices from these data, were calculated. The respective e-vals are shown in figure 4, top row of exhibits. For each set we got $d = 15$ eigenvalues denoted $\lambda_1, \ldots, \lambda_{15}$ appearing in decreasing order. The bottom exhibits of figure 4 show zoomed area of top exhibits for $y=[0, 1]$, meaning by ‘y’ the vertical axis of the plots.

![Figure 4](image)

**Figure 4** Top row: Eigenvalues obtained from correlation matrices of sets A, B and C, bottom row: zoom of top row (for values of the y-axis within the range=[0,1])

According to Properties 2, 3, and 4 introduced in Section 3, the obtained e-vals play an essential role in determining the dimensionality of the analyzed data. Looking at figure 4, one may see that except the first two or three e-vals, the remaining ones have values smaller than 1.0. Thus, according Kaiser’s principle [17], the intrinsic dimensionality of the sets is at most 3.

Set A (‘bad condition data’) has simple structure with first dominant PC and next two acceptable according to mentioned Kaiser’s rule. Set B (‘good condition data’) has the most complicated structure: its first five e-vals $\lambda_1, \ldots, \lambda_5$ are equal to 6.11, 2.45, 1.02, 0.98 and 0.64, respectively (see Table 1).

| PC | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|----|----|----|----|----|----|----|----|----|----|----|
| Set A | 9.81 | 1.07 | 1.04 | 0.63 | 0.48 | 0.37 | 0.29 | 0.28 | 0.26 | 0.20 |
| Set B | 6.11 | 2.45 | 1.02 | 0.98 | 0.64 | 0.58 | 0.52 | 0.48 | 0.45 | 0.42 |
| Set C | 10.73 | 1.66 | 0.51 | 0.43 | 0.36 | 0.25 | 0.19 | 0.17 | 0.15 | 0.14 |
It is important to note that the set C, with all data taken together, has the simplest structure. Taking data together caused a mixing of properties of data sets A and B and this data fusion has smoothed over some of the individual characteristics of the two sets. We claim that some important information (i.e. different structure of the data that is associated with different conditions of the machines) has been cancelled out and one of our conclusions is to analyse data for different condition separately.

Looking at figure 4, in particular the top row, one may see the big difference in the magnitude of the first and second eigen-values. Both of them constitute the major part of the *inertia of the data*. Set B, representing device in good state shows a prevalence of the first two eigen-values; they are dominant, however their dominance, especially for the second eigen-value, is not very big. According to Kaiser’s principle, both of them are relevant for expressing the structure of the ‘good’ data. Looking at the subplot for set A one states immediately that here only the first PC is decidedly dominant and all the others are secondary, in the background.

### 4.2. Analysis of cumulative sums of eigenvalues.

The contributions of subsequent e-vals to the *total inertia* of the data are shown in Table 2. These contributions were obtained as cumulative sums of successive rescaled (normalized) e-vals (in order to get normalised value of e-val, each e-val was divided by d=15 which is the number of all eigenvalues obtained from a correlation matrix, see Property 1 in Section 3). At the same time, see Property 2 and 3 from Section 3, the numbers presented in Table 2, may be considered as fractions of *total inertia* explained (or accounted for) by the first K (K=1, 2, … , 10) PCs.

The cumulated values of the first 10 rescaled e-vals for sets A, B, and C are shown in Table 2 below. Take notice, how differently sets A, B, and C reproduce the *total inertia* for K=2 and K=3.

#### Table 2. Cumulated eigenvalues (as fractions of Total=1.0).

| Number of PCs, K | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
|------------------|------|------|------|------|------|------|------|------|------|------|
| Set A            | 0.65 | 0.73 | 0.79 | 0.84 | 0.87 | 0.89 | 0.91 | 0.93 | 0.95 | 0.96 |
| Set B            | 0.41 | 0.57 | 0.64 | 0.70 | 0.75 | 0.78 | 0.82 | 0.85 | 0.88 | 0.91 |
| Set C            | 0.72 | 0.83 | 0.86 | 0.89 | 0.91 | 0.93 | 0.94 | 0.95 | 0.96 | 0.97 |

Different properties of data sets may be a basis for a new method of diagnostics. Using multi dimensional analysis of spectral features (variables PP1,…,PPd) it is possible to find out whether variables are seriously correlated or not. If they are correlated, one may exploit PCA to reduce dimension of data (2, 3 PCs are enough to describe properties of data set). On the contrary, for undamaged condition, there is a need to use more principal components.

### 4.3 Projection of data vectors 3D (visualisation of first 3 PCs).

It has been found in section 4.1 that real dimension of data is 3. By selection of the 3 first PCs it is possible to analyze the data geometry in 3D space. This is depicted in figure 5, which shows some snapshots when rotating the 3D configuration. It is clear that data points presented as red “+” are much more dispersed in 3D space than data points marked as green “o” (much less dispersed). One may say that system in good condition is more stable. Notice also numerous outliers, which need/have separate elaboration [18].

On the opposite, machine in bad condition works in unstable way, produces signals that provide higher variation of features, so there is a need to analyse data using statistical approach; using just a few (one) signal for such high dispersion of features may provide wrong diagnosis.
4.4 Investigation of the correlations between features in sets A and B.
In section 4.1 it has been found that real dimension of 15D data is at most 3. It means that our features are correlated and they are redundant. One may ask: what features are correlated to other and what is value of this correlation? Are these features correlated in the same way for bad and for good condition? To answer for such questions we have calculated the correlation matrices for features. Correlation matrix is of size 15x15, which yields 225 correlation coefficients $r(i,j)$, $i=1, \ldots, 15; j=1,\ldots, 15$, satisfying the inequality $-1 \leq r(i,j) \leq 1$. The elements $r(i,i)$ on the diagonal are all equal 1 (see Fig 6). Notice that in set A pairs $(i,j)$, $i,j=14,11,9,4,12,1,5,7,8,13$ are highly correlated (red colour approaching black). In set B such high correlations appear only for pairs $(5,1), (7,1), (1,4), (1,2)$ and their reverse counterparts $r(j,i)$.
It is easy to notice in figure 6 that correlation between variables is different for good and bad condition. In order to get better (more “global”) comparison we propose to analyse the histogram of all the 225 correlation coefficients encountered in set A and set B, which is shown in Fig.7. One may notice the big difference between the two displays. The histogram for set A (Bad) shows a prevalence of correlation values greater than 0.5, while for the set B (Good) the opposite is true.

**Figure 6** Visualisation of regrouped correlation coefficients obtained from set A (left panel) and set B (right panel).

**Figure 7.** Magnitude of the values \( r(i,j) \) found in correlation matrices computed from the set A (‘bad’, left exhibit) and B (‘good’, right exhibit) respectively (X axis – correlation value \( r(i,j) \), Y axis- number of correlation pairs). Notice the prevalence of high correlations in set A (left exhibit) and low correlations in set B (right exhibit).

Figure 8 provides an illustration on the differences of correlation coefficients in set A and set B. We have chosen the correlations of the variable #5, which plays a central role in the left exhibit of figure 6. Using the bootstrap method, the bootstrap samples of length \( N=5000 \) were created for the following pairs: (5,12), (5,1), (5,7), and (5,8). Histograms of the respective correlation coefficients, calculated from the generated bootstrap samples, are displayed in Fig. 8. The top row of exhibits shows the results for the good data (set B), and the bottom row the corresponding exhibits for the bad data (set A). Looking at the exhibits it becomes immediately apparent that the domain of histograms from set B and set A is completely different.
Figure 8. Distribution of correlation coefficients $r(5,12)$, $r(5,1)$, $r(5,7)$, $r(5,8)$ evaluated from $N=5000$ bootstrap samples. Top row: from set B (bad), bottom row: from set A (good). Obviously, the distributions for the respective pairs (from A and B) are quite different; their domains are distinct.

5. Discussions and Conclusions

It has been shown, that gearbox in different technical condition produces vibration signal with different spectral structure. Obviously, it is well known in general, however the analysis in this paper has been done in more advanced way. It has been found earlier that, due to damage evolution, total energy of dissipation increases, however, we don’t use such “global” measure, this paper has been focused on multidimensional data analysis in order to investigate how particular components (amplitudes of mesh frequencies) behave when damage appears, more over, we analyzed it for more difficult case – when machine was working under non-stationary operations.

As an input data we have used information extracted from spectrum of segmented vibration signal, namely 15 amplitudes of components (related to mesh frequency and its harmonics) have been acquired from each segment. These variables (called in the paper PPs) were processed using PCA. It has been shown in [12] that gearbox in bad condition generates signal with many harmonics with relatively high signal to noise ratio (SNR) and in fact all are sensitive to damage evolution and load variation. We have found here (see section 4) that our variables (PPs) are highly correlated and therefore need smaller number of Principal Components (PCs). On the contrary, gearbox in good condition produced vibration signal with smaller number of harmonics with suitable SNR, and moreover, they are less sensitive to load variation than for worn gearbox, so finally we need more PCs to describe this data sets.

Differences in obtained results for data set A (bad condition) and B (good condition) give a basis to define a new approach for diagnosis – using PCA one may find intrinsic dimension of data and make some conclusions regarding machine condition. Different properties of bad and good condition data sets have been noticed using several approaches. Our data (PPs) from “bad” and “good” data sets have: a) different intrinsic dimension (different number of PCs are required to describe data set, see Table 1), b) different distribution of Total Intertia (see Table 2), c) different structure in the domain of first three Principal Components (features from gearbox in good condition is more concentrated/less dispersed than gearbox in bad condition, figure 5) and finally d) variables (PPs) are correlated to each other with different intensity (bad condition – serious correlation, good condition – smaller correlation, see figures 6–8).

It is necessary to analyse good/bad condition data separately, taking data together to data set C has provided different properties of PCA results and cancel out information associated with different conditions of gearbox (see Tables 1 and 2)
Acknowledgments

Paper is financially supported by State Committee for Scientific Research in 2009-2012 as research project (R. Zimroz).
The authors thank to one anonymous referee for constructive comments and remarks.

References

[1] Liao G, Liu S, Shi T., and Zhang G 2004 Gearbox condition monitoring using self-organizing feature maps *Journal of Mechanical Engineering Science* 218(1) 119-129
[2] Cempel Cz and Tabaszewski M 2007 Multidimensional condition monitoring of machines in non-stationary operation *Mechanical Systems and Signal Processing* 21(3) 1233-1241
[3] Cempel Cz 2003 Multidimensional condition monitoring of mechanical systems in operation *Mechanical Systems and Signal Processing* 17(6) 1291-1303
[4] Chongqing Hu, Aihua Li and Xingyang Zhao 2011 Multivariate statistical analysis strategy for multiple misfire detection in internal combustion engines *Mechanical Systems and Signal Processing* 25 694-703
[5] Dybala J and Radkowski S 2007 Geometrical method of selection of features of diagnostic signals *Mechanical Systems and Signal Processing* 21(2) 761-779
[6] Baydar N, Chen Q, Ball A and Kruger U 2001 Detection of incipient tooth defect in helical gears using multivariate statistics *Mechanical Systems and Signal Processing*, 15(2) 303-321
[7] Trendafilova I 2010 An automated procedure for detection and identification of ball bearing damage using multivariate statistics and pattern recognition *Mechanical Systems and Signal Processing* 24(6) 1858-1869
[8] Staszewski W J 1998 Wavelet Data Compression and Feature Selection for Vibration Analysis *Journal of Sound and Vibration* 211(5) 735-760
[9] Samuel P D and Pines D J 2005 A review of vibration-based techniques for helicopter transmission diagnostics *Journal of Sound and Vibration* 282(1-2) 475-508
[10] Staszewski W J and Worden K 1997 Classification of Faults in Gearboxes, Pre-Processing Algorithms and Neural Networks *Neural Computing and Applications* 5 160-183
[11] Staszewski W J Worden K and Tomlinson G R 1997 Time-Frequency Analysis in Gearbox Fault Detection Using Wigner-Ville Distribution and Pattern Recognition *Mechanical Systems and Signal Processing* 11(5) 673-692
[12] Bartelmus W and Zimroz R 2009 A new feature for monitoring the condition of gearboxes in nonstationary operating systems *Mechanical Systems and Signal Processing* 23(5) 1528-1534
[13] Bartelmus W and Zimroz R 2009 Vibration condition monitoring of planetary gearbox under varying external load *Mechanical Systems and Signal Processing* 23(1) 246-257
[14] He Q, Yan R, Kong F and Du R 2009 Machine condition monitoring using Principal Component representations *Mechanical Systems and Signal Processing* 23/2 446-466
[15] Bellino A, Fasanoa A, Garibaldi L and Marchesiello S 2010 PCA-based detection of damage in time-varying systems *Mechanical Systems and Signal Processing* 24(7) 2250-2260
[16] Gelman L, Zimroz R, Birkel J, Leigh-Firbank H, Simms D, Waterland B and Whitehurst G 2005 Adaptive vibration condition monitoring technology for local tooth damage in gearboxes. *Insight International Journal of Non-Destructive Testing and Condition Monitoring* 47(8) 461-464
[17] Jolliffe IT 2002 Principal Component Analysis (New York: Springer)
[18] Bartkowiak A and Zimroz R 2011 Outliers analysis and one class classification approach for planetary gearbox diagnosis (DAMAS Conference 2011, Oxford, under review)
[19] Zimroz R and Bartkowiak A 2011 Two simple multivariate procedures for monitoring planetary gearboxes working in non-stationary operating conditions (Submitted to *Mechanical Systems and Signal Processing*, Aug 2010, under revision)
[20] Wise B M The corrmat function *Eigenvector Research Inc.*
(http://www.eigenvector.com/MATLAB/corrmap.html)