STATISTICS IN TRANSITION new series, March 2019
Vol. 20, No. 1, pp. 131–151, DOI 10.21307/stattrans-2019-008

# AN APPLICATION OF FUNCTIONAL DATA ANALYSIS TO LOCAL DAMAGE DETECTION

**Jacek Leśkow**[1], **Maria Skupień**[2]

## ABSTRACT

Vibration signals sampled with a high frequency constitute a basic source of information about machine behaviour. Few minutes of signal observations easily translate into several millions of data points to be processed with the purpose of the damage detection. Big dimensionality of data sets creates serious difficulties with detection of frequencies specific for a particular local damage. In view of that, traditional spectral analysis tools like spectrograms should be improved to efficiently identify the frequency bands where the impulsivity is most marked (the so-called *informative frequency bands* or IFB). We propose the functional approach known in modern time series analysis to overcome these difficulties. We will process data sets as collections of random functions to apply techniques of the functional data analysis. As a result, we will be able to represent massive data sets through few real-valued functions and corresponding parameters, which are the eigenfunctions and eigenvalues of the covariance operator describing the signal. We will also propose a new technique based on the bootstrap resampling to choose the optimal dimension in representing big data sets that we process. Using real data generated by a gearbox and a wheel bearings we will show how these techniques work in practice.

**Key words:** damage detection, functional data, functional principal components, informative frequency band.

## 1. Introduction

In recent years, extensive research has been focused on big data problems related to statistical signal processing. The big data problem arises when a structural health monitoring system is supported by on-line sensors producing a signal observed with e.g. 20 kHz frequency. After several hours of observations we have millions of data points that can be used for processing. So far, many practical applications have been based on selecting some segments of data and classical analyses have then been conducted on selected segments. However, modern statistical inference can be based on the whole multi-million points sample when the functional data analysis approach is used (see, for example Horváth and Kokoszka, 2012). This is especially suitable when we deal with time-varying systems and when techniques related to time-frequency analysis are used. For example, in Yang and Nagarajaiah (2014), results are presented on independent component analysis with

[1]Cracow University of Technology, Poland. E-mail: jleskow@pk.edu.pl.
[2]Pedagogical University of Cracow, Poland. E-mail: marysia.skupien@gmail.com. ORCID ID: https://orcid.org/0000-0003-1480-0810.

wavelet transform. Interesting exploratory studies on frequency response function (FRF) can be found in Staszewski and Wallace (2014).

In the case of big data and time varying systems, it is convenient to consider data as curves. For example, for a rotating element the natural data curve would be generated in the interval of the length of a cycle. For other signals, a natural interval may be a second or a minute. From this perspective, the spectrogram (see, e.g. Gryllias, et al., 2017 or Khadersab and Shivakumar, 2018) is a way of converting a big signal into segments via Fourier analysis on sliding blocks creating a time-frequency map. Such a map is exactly a collection of random curves. From this point of view, a multimillion data points segment of a signal generated by a sensor attached to some structure is seen as a collection of curves.

In recent years there is a significant research dedicated to wheel bearing diagnostics. Randall and Antoni (2011) presents a review of contemporary techniques. Other publications like (Liu, et al., 2018) or (Jia, et al., 2016) present contemporary artificial intelligence technique and their applications in the wheel bearing diagnosis. However, according to our knowledge, so far no one has implemented modern statistical inference tools, based on functional data approach, to the diagnosis of wheel bearings. In this context we would like to mention research in Spiridonakos and Fassois (2014) dedicated to functional time series and their applications to non-stationary random vibrations, where a functional approach is proposed in a different context.

The main line of our article is to show how to use modern statistical tools like functional data analysis or bootstrap to efficiently process big data sets and identify significant frequencies. We propose a new perspective in looking at a very popular tool in signal analysis such as the spectrogram. In Subsection 3.1. we explain the difference between classical spectrogram and new functional one proposed by us. Usually, the spectrogram is generated by a signal coming from an excavating machine or by a signal generated by a wheel bearing (see, for example Cioch, et al. 2013). Then, the proper functioning of the tested system is diagnosed by identifying the frequency band where the signal impulsivity is most marked, called in the sequel informative frequency bands (IFB) (see Randall and Antoni, 2011 or Obuchowski, et al., 2014).

We propose to view the spectrogram as a collection of random curves. Using the functional data analysis approach, we are able to process such data and quickly solve the problem of identifying the IFB. The main advantage of our method is a possibility of using large data sets to generate few dimension of the diagnostic analysis. Millions of data points are represented as random curves, then in the appropriately defined infinitely dimensional Hilbert space the covariance operator is considered and its empirical counterpart is studied. Finally, only few eigenvalues and eigenfunctions of the empirical covariance operator are sufficient to efficiently represent the signal at hand. We propose a novel technique of using bootstrap resampling in deciding on dimensionality reduction. Another main advantage of our method is that no matter what frequency of signal sampling might be, our method provides a uniform result. On the diagram in Figure 1 we show the logic of our

approach.

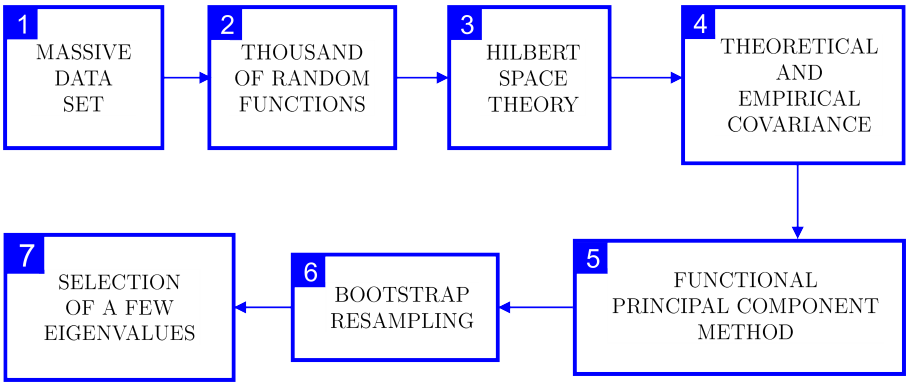



Figure 1: Flowchart of the main idea of our paper.

Our article is organized as follows. In Section 2, we present main elements of the functional data analysis approach as applied to signals. Basic tools such as functional principal components, Hilbert space valued random transformations, variance and covariance operators are presented there. In Section 3, we apply this approach to the problem of identifying the informative frequency bands for a spectrogram. In that context we show that without discarding any data points we are able to reduce the dimensionality of data to just a few eigenvalues and few eigenfunctions and retain more than 80% of its energy. Finally, in Section 4 we provide a short discussion of our results.

## 2. Functional data approach in statistical signal processing

Our starting point here is the new perspective on statistical signal processing from the functional data analysis point of view. To start, assume that we observe a signal $\{X(s) : s \in [0,T]\}$ and the collect time $T$ is really huge, e.g. in the order of several millions of individual data points. We will view such a signal as a collection of random curves $\{X_n(s), s \in [n, n+w]\}$ each defined on the interval $[n, n+w]$ with the width $w$. These curves may be considered independent or correlated, depending on the model and a context of study. For example, the spectrogram technique Gryllias, et al., (2017), Khadersab and Shivakumar, (2018) transforms a long signal $\{X(s) : s \in [0,T]\}$ into a collection $\{x_1(f), \ldots, x_N(f)\}$ of spectral densities defined on a common frequency interval $[0,\Lambda]$, where $f \in [0,\Lambda]$. In the Subsection 2.3, in the last algorithm, we explain what our functional observations are and how they were obtained from a discrete vibration signal. We assume those observations to be independent. We are aware that the technic of the overlapping window may introduce some dependence into our data structure. At this point we neglect this dependence and we proceed as if the data were independent. However, in the

literature there is a number of cases in which methods of functional data analysis have been adapted to series or signals by cutting them as if they were curves observed independently (see Ramsay and Silverman (2002) and (2005)).

To simplify our notation and with no loss of generality, we will assume that $\Lambda = 1$ so all data curves are defined on the unit interval $[0,1]$ . The observed curves will be assumed to be square integrable. A natural choice of the realization space will therefore be the Hilbert space $H = L^2[0,1]$. This is a consequence of the expansion methods (see Ramsay and Silverman (2002) and (2005)), where the functional form of curves is obtained by a linear span of base functions. The relevant coefficients are then estimated from discrete observation of curves at different time points by least squares methods. This point of view allows us to introduce a Hilbert space of squared integrable functions, where the theory of functional principal component analysis (and many other functional methods) can be applied. From this perspective our initial signal $\{X(s) : s \in [0,T]\}$ can be viewed as a collection of random curves $\{X_n\}$, each in the space $H$. For such random curves, we will now introduce concepts of mean, variance and covariance.

Note that each random curve $X$ is as a random element acting from some probability space $(\Omega, \mathscr{F}, P)$ onto $L^2[0,1]$. If $X$ is integrable, then there is a unique function $\mu \in L^2$ such that $\mathbb{E}\langle y, X\rangle = \langle y, \mu\rangle$ for each $y \in L^2$. It follows that $\mu(t) = \mathbb{E}[X(t)]$ for all $t \in [0,1]$. Here $\langle \cdot, \cdot \rangle$ is the scalar product in the Hilbert space $H$ defined as $\langle x, y\rangle = \int_0^1 x(s)y(s)ds$ with the norm defined as: $\|f\| = \sqrt{\int_0^1 f^2(t)\, dt}$ for all $f \in H$. For more mathematical details regarding statistics on Hilbert space the reader is referred to Horváth and Kokoszka (2012).

We recall here the notion of spectral decomposition for matrixes and functional operator.

**Theorem: 1.** *Suppose $\boldsymbol{A}$ is a symmetric, positive definited $k \times k$ matrix. Then, there is an orthonormal matrix $\boldsymbol{U} = [\boldsymbol{u_1}, \ldots, \boldsymbol{u_k}]$ whose columns are the eigenvectors of $\boldsymbol{A}$, i.e.*

$$\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I} \qquad \text{and} \qquad \boldsymbol{A}\boldsymbol{u_j} = \lambda_j \boldsymbol{u_j}$$

*Moreover, $\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{U} = Diag[\lambda_1, \ldots, \lambda_k]$ The orthonormality of $\boldsymbol{U}$ is equivalent to the assertion that the vectors $\boldsymbol{u_1}, \ldots, \boldsymbol{u_k}$ form an orthonormal basis in the Euclidean space $\mathbb{R}^k$. Theorem 1 implies that*

$$\underset{(k\times k)}{\boldsymbol{A}} = \sum_{i=1}^{k} \lambda_i \underset{(k\times 1)(1\times k)}{\boldsymbol{u_i}\ \boldsymbol{u_i}^T} = \underset{(k\times k)}{\boldsymbol{U}}\ \underset{(k\times k)}{\boldsymbol{\Lambda}}\ \underset{(k\times k)}{\boldsymbol{U}^T},$$

*a representation known as a spectral decomposition of $\boldsymbol{A}$.*

The above ideas can be easily extended to a separable Hilbert space. Suppose $\boldsymbol{\Psi}$ is a symmetric positive–definite Hilbert–Schmidt operator in $L^2$. Covariance operator (1) and its sample counterpart (2) are in this class, provided $\mathbb{E}\|X\|^4 < \infty$. The

operator $\boldsymbol{\Psi}$ then admits the functional counterpart of spectral decomposition (1)

$$\langle \boldsymbol{\Psi}(x), x \rangle = \left\langle \sum_{i=j}^{\infty} \lambda_j \langle x, v_j \rangle v_j, x \right\rangle = \sum_{i=j}^{\infty} \lambda_j \langle x, v_j \rangle^2$$

where scalars $\lambda_j$ are eigenvalues and $v_j$ corresponding eigenfunctions, satisfying equation $\boldsymbol{\Psi}(v_j) = \lambda_j v_j$.

### 2.1. Theoretical and empirical covariance operators

Since we are adopting the Hilbert space approach, the usual covariance will be an operator, that is a transformation from the Hilbert space to the Hilbert space. This is analogous to the traditional concept of covariance, where a real-valued signal $X$ generates a covariance function transforming real values to real values. Let us have a closer look at the formal definition of the covariance operator.

For $X$ integrable and $\mathbb{E}X = 0$, the covariance operator of $X$ is defined by

$$C(x) = \mathbb{E}[\langle X, x \rangle X], \quad x \in L^2, \tag{1}$$

where

$$\begin{aligned} C(x)(t) =& \mathbb{E}[\langle X, x \rangle X(t)] = \mathbb{E} \int_0^1 X(s)x(s)ds X(t) = \\ =& \int_0^1 \underbrace{\mathbb{E}[X(s)X(t)]}_{=c(s,t)} x(s)ds = \int_0^1 c(s,t)x(s)ds. \end{aligned}$$

In the sequel, the covariance operator $C$ will be our central point of a study as it fully describes the energy generated by the random element $X$, which in turn represents a signal under study. While studying the covariance operator, we will focus on characterizing its eigenvalues. They will be important in reducing the dimensionality of $C$ to just a few of non-negative numbers. For more theoretical properties of the covariance operators see Horváth and Kokoszka (2012).

The main task of statistical signal processing in the functional data analysis context will be to introduce an empirical covariance operator $\hat{C}$ that is fully defined by random curves $x_1, \ldots, x_N$ and for sufficiently large sample size $N$ approximates the theoretical covariance operator $C$ that describes the signal of interest. Therefore, assume that a sample of random functions $x_1, \ldots, x_N$ corresponds to the signal $X$. Recall that the spectrogram can be viewed as a collection of random curves with arguments in the frequency interval. In general, however, such random functions can represent segments of signals from different time intervals or replica of signals collected via some transformations.

For $x_1,\ldots,x_N$ we define the sample covariance operator as:

$$\hat{C}(x) = \frac{1}{N}\sum_{i=1}^{N}\langle x_i, x\rangle x_i,\ x \in H. \tag{2}$$

It is important to note that in the formula (2) the symbol $x$ corresponds to any function $x$ from the Hilbert space $H$ while $x_i$ is the *observed* random function $x_i$ generated by the signal of interest. In such a way the estimator $\hat{C}$ given in (2) approximates the theoretical covariance operator $C$ defined in (1). For more mathematical theory related to this approximation the reader is refereed to (Bosq, 2000).

For covariance operators which are symmetric, positive defined and are defined on a Hilbert space and are Hilbert-Schmidt operators we have a very interesting property. Suppose $\mathbf{\Psi}$ is a symmetric, positive definite Hilbert–Schmidt operator with eigenfunctions $v_j$ and eigenvalues $\lambda_j$, satisfying $\lambda_1 > \lambda_2 > \cdots$. Then,

$$\sup_{\|x\|=1}\{\langle \mathbf{\Psi}(x), x\rangle : \langle x, v_j\rangle = 0,\ 1 \le j \le i-1, i < p\} = \lambda_i$$

and the supremum is reached if $x = v_i$. The maximizing function $x$ is unique up to a sign. The upper bound $p$ for index $i$ is defined in Subsection 2.2.

Of course, the empirical covariance operator $\hat{C}$ given in (2) satisfies the above property. This in turn gives us the following facts fundamentally important in the subsequent statistical considerations:

- the empirical covariance operator $\hat{C}$ defined in (2) is fully defined by its eigenfunctions and eigenvalues,
- the covariance and the total variance of the sample (thus the signal) will be described by the estimated eigenvalues.

In what follows, we will show how to apply these facts. To start, recall that the random functions $x_1,\ldots,x_N$ correspond to the signal of interest. Now, fix the integer number $p \ll N$. Next, choose the basis $u_1, u_2, \ldots$ in $H$ such that:

$$\hat{S}^2 = \sum_{i=1}^{N}\left\| x_i - \sum_{k=1}^{p}\langle x_i, u_k\rangle u_k \right\|^2 \leftarrow \min$$

Then, each curve $x_i$ can be approximated by $\sum_{k=1}^{p}\langle x_i, u_k\rangle u_k = \sum_{k=1}^{p} c_k u_k$. Hence, an infinite dimensional curve $x_i$ is represented by a $p$ variate vector $(\langle x_i, u_1\rangle, \ldots, \langle x_i, u_p\rangle)$. Now we will use the fundamental fact that the basis elements $u_1,\ldots,u_p$ can be chosen to correspond to the eigenfunctions of the sample covariance $\hat{C}$. More precisely, functions $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_p$ minimizing $\hat{S}^2$ are equal (up to a sign) to normalized eigenfunctions of the sample covariance operator $\hat{C}$.

Note that scores $\langle x, \hat{u}_k\rangle = \int_0^1 x(t)\hat{u}_k(t)dt$ measure the importance of the $k$th function $\hat{u}_k$ in the representation

$$x \approx \sum_{k=1}^{p} \langle x, \hat{u}_k \rangle \hat{u}_k. \tag{3}$$

In the sequel, we will call $\hat{u}_k$ the $k-th$ *functional principal component*. One of its important properties is orthonormality.

The eigenvalues are extremely important in describing the total energy of the signal. For a random element $X$ with values in the Hilbert space $H$ we have:

$$\mathbb{E}\|X\|^2 = \sum_{j=1}^{\infty} \mathbb{E}\langle X, v_j \rangle^2 = \sum_{j=1}^{\infty} \langle C(v_j), v_j \rangle = \sum_{j=1}^{\infty} \lambda_j.$$

The quantity $\mathbb{E}\|X\|^2$ can be called *theoretical total variance*. Its empirical equivalent, *sample total variance*, based on a sample of random functions $x_1, \ldots, x_N$ is defined as:

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\|x_i\|^2 &= \frac{1}{N}\sum_{i=1}^{N}\langle x_i, x_i\rangle = \frac{1}{N}\sum_{i=1}^{N}\Big\langle \sum_{j=1}^{N}\langle x_i, \hat{u}_j\rangle \hat{u}_j, \sum_{j=1}^{N}\langle x_i, \hat{u}_j\rangle \hat{u}_j \Big\rangle = \\ &= \sum_{j=1}^{N}\frac{1}{N}\sum_{i=1}^{N}\langle x_i, \hat{u}_j\rangle^2 = \sum_{j=1}^{N}\langle \hat{C}(\hat{u}_j), \hat{u}_j\rangle = \sum_{j=1}^{N}\sum_{k=1}^{\infty}\hat{\lambda}_k \underbrace{\langle \hat{u}_j, \hat{u}_k\rangle^2}_{=\delta_{jk}} = \sum_{j=1}^{N}\hat{\lambda}_j,\end{aligned}$$

where $\delta_{jk}$ is the Kronecker delta - a function of two variables, defined as follows: $\delta_{jk} = \delta(j,k) = \begin{cases} 1, & \text{if} \quad k = j \\ 0, & \text{if} \quad k \neq j \end{cases}$ and $\hat{\lambda}_j$ is interpreted as variance in the direction $\hat{u}_j$. In other words, the empirical functional principal component $\hat{u}_j$ generated by the empirical covariance operator $\hat{C}$ explains the fraction of the total sample variance equal to $\hat{\lambda}_j / \sum_{k=1}^{N} \hat{\lambda}_k$. The above approach will be referred to as the Functional Principal Component Analysis or FPCA for short.

### 2.2. Reduction of dimensionality

While working with big data sets generated by signals, the crucial point is to select the number $p$ of eigenvalues that give a reasonable approximation of the sample total variance. One of the methods of selecting $p$, for which function $x_i$ has the best approximation given by the formula $\sum_{j=1}^{p} \langle x_i, \hat{u}_j \rangle \hat{u}_j$ is the CPV method. This method is based on calculating the cumulative percentage of the total variance

(CPV) explained by the first $p$ empirical functional principal components

$$CPV(p) = \frac{\sum_{i=1}^{p} \hat{\lambda}_i}{\sum_{i=1}^{N} \hat{\lambda}_i}. \tag{4}$$

We choose $p$ for which $CPV(p)$ exceeds a desired level. Ideally, one would like to recover 100% of the total variance, however in practical situations we usually settle with 80% or higher. Such approach has a dramatic effect on our ability to process big data generated by the signals observed with high frequency over long periods of time. First, we split the signal into a sequence of random functions and then we follow the approach above to identify the first $p$ eigenvalues. In the following section we will show that for vibration data coming from the gearbox of the excavating machine choosing $p$ as small as 8 retrieves a large percentage of the total variance.
Below, we present our original method, based on bootstrap technique, which allows us to precisely evaluate the percentage of variance explained with a confidence interval.

**CPV bootstrap algorithm.**

**Step 1**. We start from the initial sample of random functions $x_1,\ldots,x_N$. We sample with replacement the first bootstrap sample $x_1^{*1},\ldots,x_N^{*1}$ from the initial set $x_1,\ldots,x_N$. It is important that the bootstrap sample is of the same size as the original one. For such bootstrap sample we calculate the first bootstrap value $CPV(p)^{*1}$ of $CPV(p)$ (see formula (4)).

**Step 2**. We repeat Step 1 B times. Usually, we take B = 1000. As a result, we get B bootstrap replications $\{CPV(p)^{*1},\ldots,CPV(p)^{*B}\}$.

**Step 3**. We produce a 95% confidence interval for $CPV(p)$ using 2.5% and 97.5% empirical quantiles from the replications $\{CPV(p)^{*1},\ldots,CPV(p)^{*B}\}$.

We illustrate the logic of our bootstrap procedure on the diagram in Figure 2.
The below procedure is admissible from the statistical point of view as it is reconstructing the true unknown distribution of the $CPV(p)$, which in turn is based on the unknown distribution of eigenvalues. For a more detailed discussion related to eigenvalues distribution in the functional approach see e.g. (Mas, 2002). We would like to emphasize that our method allows us to analyse big data sets generated by signals observed over a long period of time using just 8 eigenvalues and 8 associated eigenfunctions. In general, one can start with even 2 eigenvalues, calculate the $CPV(3)/CPV(2)$ and bootstrap it to get its confidence intervals and then see whether adding third eigenvalue significantly improves CPV(3) as compared to

CPV(2). The reader can find an explanation of this improvement in Subsection 3.2 on the example of analysed data set. Additional argument is provided by studying the convergence of the ratio $CPV(p+1)/CPV(p)$ and identifying the proper $p$, where it starts to stabilize (Figure 3).

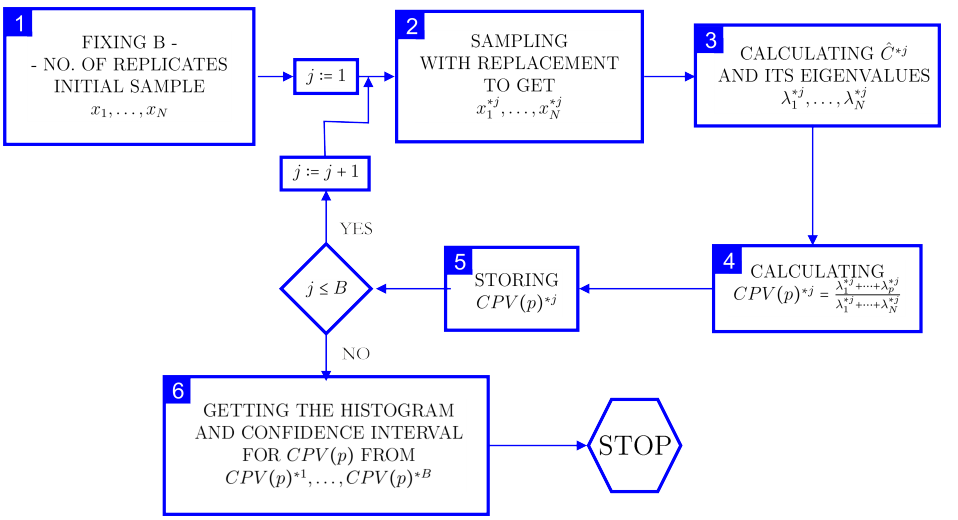



Figure 2: Flowchart of finding the empirical distribution of $CPV(p)$ via bootstrap algorithm.

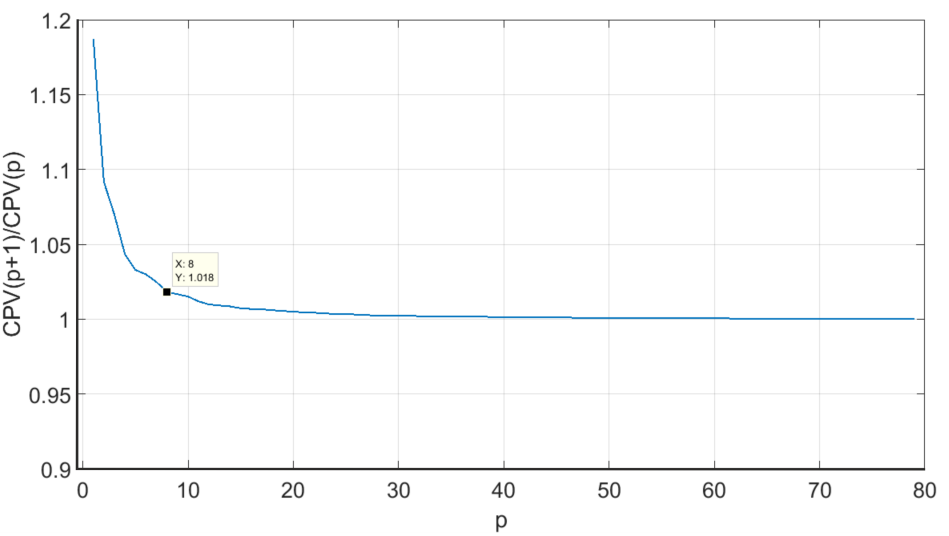



Figure 3: Visualization of the ratio $CPV(p+1)/CPV(p)$ based on experimental data set.

In the following subsection, we will show how to apply the dimensionality reduc-

tion obtained from FPCA to the problem of identification of informative frequency bands (IFB).

### 2.3. FPCA and informative frequency bands for spectrogram

Let us start with a definition. We define the Informative Frequency Band (IFB) generated by a series of squared absolute of Short-Time Fourier Transforms $|STFT(t,f)|^2$, $t=0,\ldots,T$ and $f\in[0,\Lambda]$ as such a subset $A\subset[0,\Lambda]$ that

$$\frac{IE}{E}(A)\stackrel{def}{=}\frac{\sum\limits_{f\in A}\sum\limits_{t=0}^{T}|STFT(t,f)|^2}{\sum\limits_{f\in[0,\Lambda]}\sum\limits_{t=0}^{T}|STFT(t,f)|^2}\geq L, \tag{5}$$

where $E$ is the total energy of the signal and $IE$ is the energy within the frequency set $A$ and $STFT(t,f)$ defined in (6). The threshold value $L$, $0\leq L\leq 1$ is usually selected to be bigger than 80%.

To simplify the search for the frequency set $A$ defined in (5), we start from a one-element subset that contains the most energy and we augment it successively adding element by element in the order of the energy contribution. This is represented by the algorithm below.

**Identification of IFB by STFT.**

**Step 1.** Identify first $f_1$ such that $\sum_{t=0}^{T}|STFT(t,f_1)|^2$ is the biggest.
**Step 2.** Obtain the ranking of frequencies $f_i$ via

$$\sum_{t=0}^{T}|STFT(t,f_1)|^2\geq\ldots\sum_{t=0}^{T}|STFT(t,f_i)|^2\geq\ldots\sum_{t=0}^{T}|STFT(t,f_I)|^2\ .$$

where $I$ is the cardinality of a set of frequencies (discretizated interval of frequencies).
**Step 3.** From the above ranking we identify the subset $A_{spectr}=\{f_1,f_2,\ldots,f_K\}$ induced by spectrogram such that

$$\frac{IE}{E}(A_{spectr})\geq L,$$

where $\frac{IE}{E}(\cdot)$ was defined in (5).

The above procedure does not involve FPCA, it simply ranks the frequencies in the decreasing order of their influence on the total variability generated by STFT. We will now show how FPCA and the reduction of dimensionality obtained via $CPV(p)$ helps identify the impulsive frequencies.

**IFB identification algorithm via FPCA.**

**Step 1.** Represent the raw data produced by the spectrogram as random functions $x_1,\ldots,x_N$. In the Subsection 3.1, we explain how to obtain curves $\{x_i,\ldots,x_N\}$. Using the functional approach, find the empirical covariance estimator $\hat{C}$ (see (2)), the corresponding eigenvalues $\{\hat{\lambda}_i, i=1,\ldots,N\}$ and the functional principal components $\{\hat{u}_i(f), i=1,\ldots,N;\ f\in[0,\Lambda]\}$ (see (3)). Using the $CPV(p)$ technique identify your choice of $p$. We still use formula (5) with one modification - in place of $|STFT(t,f)|^2$ we insert FPC representation of the curve, i.e. $|x_t(f)|^2 = |\sum_{j=1}^{p}\langle\hat{u}_j,x_t\rangle\hat{u}_j(f)|^2$

**Step 2.** Identify first $f_1$ such that $\sum_{t=0}^{T}|x_t(f_1)|^2$ is the biggest.

**Step 3.** Obtain the ranking of frequencies $f_i$ via

$$\sum_{t=0}^{T}|x_t(f_1)|^2 \geq \sum_{t=0}^{T}|x_t(f_2)|^2 \geq \ldots \geq \sum_{t=0}^{T}|x_t(f_I)|^2$$

**Step 4.** From the above ranking we identify the FPCA induced subset $A_{FPCA} = \{f_1,f_2,\ldots,f_R\}$ such that

$$\frac{IE}{E}(A_{FPCA}) \geq L,$$

where $\frac{IE}{E}(\cdot)$ was defined in (5).

In the next Section dedicated to applications we will show how close the sets $A_{spectr}$ and $A_{FPCA}$ are.

## 3. Application to gearbox and wheel bearing data

In this Section we will show the application of the functional data approach presented in the previous section to the spectrogram in the context of identifying the informative frequency band (IFB). Recall that the spectrogram represents the signal as a collection of short-time Fourier transforms (STFT). Using time-frequency plots produced by sequences of STFTs one tries to identify the frequency bands where the excitation (energy) of the signal of interest is the most significant. From our perspective, however, the spectrogram is a collection of random curves. To make our point more precise let $X$ be a signal of interest. Recall that the STFT is defined as:

$$\text{STFT}(t,f) = \int_{-\infty}^{\infty} w(t-\tau)X(\tau)e^{-2\pi i f\tau}d\tau, \tag{6}$$

where $w(\cdot)$ is the window function, $t\in[0,T]$ and the frequency $f\in[0,\Lambda]$. The discrete version of STFT is defined as follows:

$$\mathrm{STFT}(t,f) = \sum_{k=0}^{M-1} X_k w(t-k) e^{-2\pi i f k/M}. \tag{7}$$

### 3.1. Spectrogram and FPCA on data

A spectrogram is a visual representation of the spectrum of frequencies in a signal as it varies with time or some other variable. A common format is a graph with two geometric dimensions: the vertical axis represents frequency, the horizontal axis is time; a third dimension indicating the amplitude of a particular frequency at a particular time is represented by the intensity or colour of each point in the image. Here we will analyse the spectrogram generated by the open-pit excavating machine from a Polish brown coal mine. The signal is the acceleration signal generated in the gearbox of this machine. The raw signal that combines four signals from four sensors has a length of 20480 data points. The impulses are convoluted with the raw signal and then the spectrogram of the result is analysed. The resulting signal $\{X(s) : s \in [0,T]\}$ is represented in Figure 4.

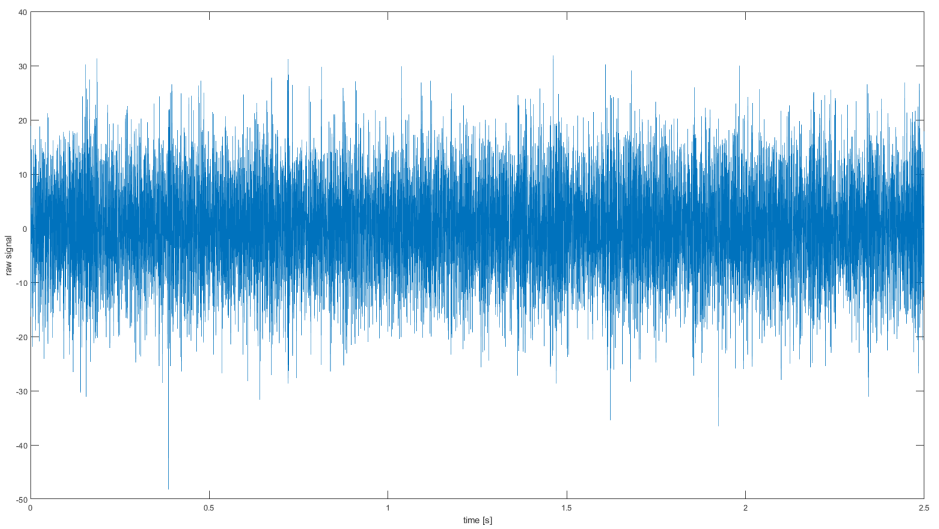



Figure 4: Acceleration signal with the impulses.

The spectrogram corresponding to the above signal is represented in Figure 5 on the left panel. The spectrogram of the signal obtained after the dimensionality reduction done by FPCA method is shown on the right panel. Here, we clarify how to create particular spectrograms.

**Traditional data spectrogram:** First, decompose the signal $\{X(s) : s \in [0,T]\}$ into the set of overlap narrowband sub-signals $\{X_t(s) : s \in [t,t+w]\}_{t=0}^{N-1}$. Next, use Fourier transform (FFT) to calculate the magnitude of the frequency spectrum for each sub-signal (FFT is a digital process). Vertical (or horizontal) line in the image cor-

responds to each sub-signals; a measurement of magnitude vs. frequency for a specific moment in time. Finally, these spectrums or time plots are then "laid side by side" to form the image or a three-dimensional surface, or slightly overlapped (windowing) in various ways. To sum up, a spectrogram is a frequency-time domain map, representing an energy of signal (power spectrum). Its values are stored in a huge matrix $\{|\text{STFT}(t,f)|^2\}_{f\in[0,\Lambda];\, t\in[0,T]}$ with entries defined in (7).

**Spectrogram via FPCA method:** We process a vibration signal similarly to traditional method with the difference that each sub-signal selected by windowing is converted into function, hence we view a spectrogram as a collection of random curves. According to our notation, for each $t = 0,\ldots,N-1$ we convert $\{|STFT(t,f)|^2\}_{f=0}^{\Lambda}$ into random curves $\{x_t(f) : f \in [0,\Lambda]\}_{t=0,\ldots,N-1}$, where $\Lambda$ is fixed maximum frequency, and $x_t \in L^2[0,\Lambda]$. Conversion of vectors (here, rows of the matrix $\{|STFT(t,f)|^2\}_{f=0}^{\Lambda}$) to curves is carried out with the use of basis expansion in $L^2[0,1]$ space. Here, we used the program `R` and its package `fda.usc` with the function `fdata2fd` to produce functional object. The type of functional basis is Bspline by default, but of course it is possible to change basis to any other, for example Fourier basis. Next, we reduce dimensionality of those curves using FPCA. For curves, represented by combination of several eigenfunctions and scores we calculate STFT defined in (6) achieving again a matrix $\{|\text{STFT}(t,f)|^2\}_{f\in[0,\Lambda];\, t\in[0,T]}$ which has graph representation in Figure 5 on right panel. From our perspective, the spectrogram is a collection of random curves indexed by the parameter $t$ (time). In other words, the spectrogram is a set which define the energy of functions. Consequently, IFB is a subset of those functions whose energy within this band is close to the total energy.

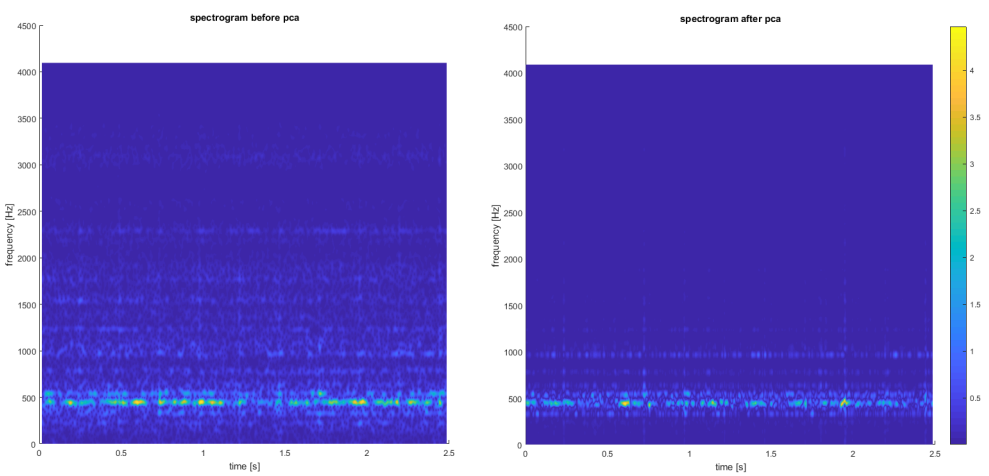



Figure 5: Traditional data spectrogram (left) and spectrogram via FPCA method (right).

Our data set is a relatively big $513 \times 1265$ matrix with entries $\{|STFT(t,f)|^2\}$. The second dimension (frequencies $f$) range is $513 = 2^9 + 1$ and it refers to the sample

frequency. The first dimension (time $t$) is $1265$ and it refers to the number of windows $N$ covering a signal of length 20480 ($16 \cdot N = 20480 - 256$). Here, we have used Hamming window (a function describing the way of sampling within a signal, given by formula $w(t-k) = 0.53836 - 0.46164\cos(\frac{2\pi(t-k)}{M-1})$; $k = 1,\ldots,M$) of length $256 = 2^8$ and an overlap $M = 240$. Next, each row of $\{|STFT(t,f)|^2\}$ is converted to function $x_t(f)$, $t = 0,\ldots,1-N$ , which forms our functional data set. The last step is to approximate those curves with principal components expansion.

We will now show how the FPCA method works in practice. We choose $p = 8$ first functional principal components generated by the empirical covariance operator $\hat{C}$ induced by the random curves $x_1,\ldots,x_N$ corresponding to the spectrogram. How to obtain the empirical operator was explained in previous Section 2, i.e. in formula (2). It is worth to notice, that $x$ in (2) is any element of the Hilbert space $H$, so the easiest way to choose this element is to take any basis element. Moreover, $\hat{C}$ is fully described by its eigenfunction and eigenvalues, which are known from the data. We illustrate our eight empirical principal components in Figure 6, which is a fragment of a 8×8 matrix of small pictures. On the diagonal, first four from the eight functional principal components $\hat{u}_i$, $i = 1,\ldots,8$ are represented. Outside the diagonal, we show the scatterplots of scores $\{(PC_i^k, PC_j^k)\}_{k=1}^N$ , $i \neq j$, $i,j = 1,\ldots,8$. Recall that scores $PC_i^k = \langle \hat{u}_i, x_k \rangle$, where $\hat{u}_i$ are the eigenfunctions of the empirical covariance operator $\hat{C}$ and $x_k$, $k = 1,\ldots,N$ is a function from the sample $\{x_1,\ldots,x_N\}$. In our case $N = 1265$. The more the scatterplots are irregular, the less correlation between them. Ideally, one would like to have a zero correlation between them since we want our functional principal components to be orthogonal. Our graphs confirm the idea of the weak correlation between the scores, hence our empirical FPC are indeed orthogonal.

In Table 1, we show the percentage of variance explained by each of the first eight functional principal components $FPC(i), i = 1,\ldots,8$.

Table 1: Percentage of variability explained by each empirical functional component.

| FPC1 | FPC2 | FPC3 | FPC4 | FPC5 | FPC6 | FPC7 | FPC8 |
|---|---|---|---|---|---|---|---|
| 52.62% | 9.91% | 5.48% | 4.07% | 3.36% | 2.38% | 2.10% | 1.96% |
| Total explained: | 81.88% | | | | | | |

First four empirical functional components are quite informative, explaining as much as 72.08% of variability so we show the corresponding 4×4 matrix of plots of the scores, which are highly uncorrelated confirming orthogonality of empirical FPC (see values of correlation in the top of the boxes outside the diagonal).

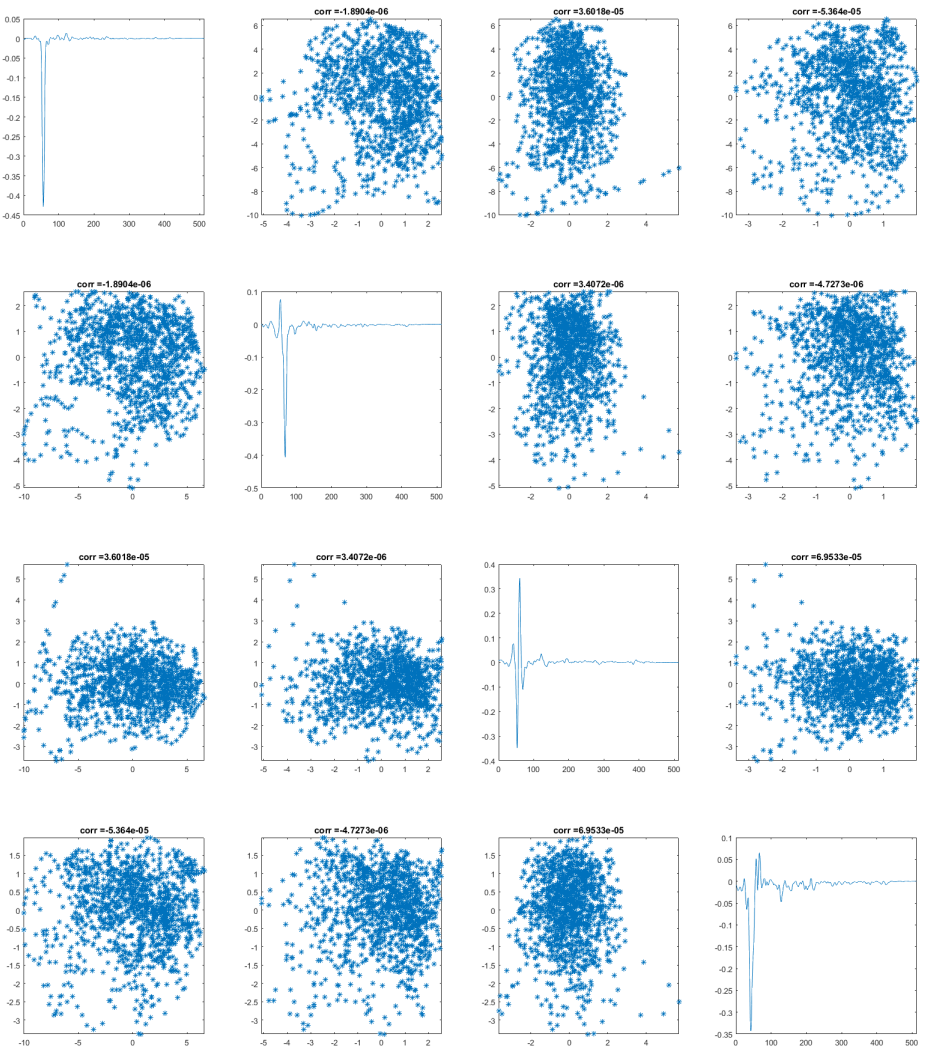
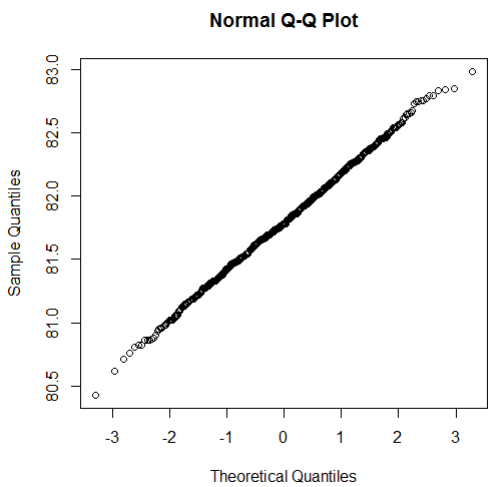



Figure 6: First four FPC with scatterplots of scores.

### 3.2. Cumulative percentage of variance (CPV) study

The previous subsection was devoted to illustrating how our FPCA method works in practice. For the first eight empirical functional eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_8$ we have obtained a quite reassuring result: they represent as much as 81.88% of variance. From the statistical perspective, however, we would like to get more information on the variability of the $CPV(p)$ coefficient. In other words, we would like to be able to measure the variability of our estimate with the point value of 81.88%. To answer this question, we will apply the CPV bootstrap algorithm introduced in the previous Section 2.
The statistical features of the bootstap distribution of CPV(8) are shown in Figure 7.

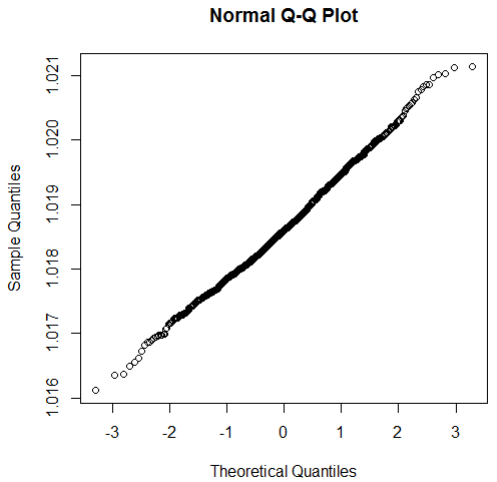



| Descriptive statistics of CPV(8) | Values |
|---|---|
| Minimum | 80.43000 |
| 2.5% quantile | 81.02000 |
| Median | 81.78000 |
| Mean | 81.79348 |
| 97.5% quantile | 82.54025 |
| Maximum | 82.98000 |
| Variance | 0.14648 |
| Skewness | −0.00285 |
| Kurtosis | 3.07514 |

Figure 7: Descriptive statistics for $CPV(8)$ based on bootstrap samples.

The most important message from the above calculations is that the 95% confidence interval for CPV(8) is from 81.02% (the 2.5% quantile) to 82.54% (the 97.5% quantile). This means that CPV as a random variable is quite concentrated around its point value 81.88% and that our results are quite reliable and have only a small spread.

To analyse the speed of convergence $CPV(p+1)/CPV(p) \stackrel{p\to\infty}{\longrightarrow} 1$ we have performed the bootstrap distribution study of this ratio and have obtained the results presented in Figure 8.




| Descriptive statistics of CPV(9)/CPV(8) | Values |
|---|---|
| Minimum | 1.01612 |
| 2.5% quantile | 1.01719 |
| Median | 1.01859 |
| Mean | 1.01864 |
| 97.5% quantile | 1.02023 |
| Maximum | 1.02113 |
| Variance | $6.56e-07$ |
| Skewness | 0.19956 |
| Kurtosis | 2.95591 |

Figure 8: Normality of $CPV(9)/CPV(8)$ based on bootstrap samples.

Again, we see the usefulness of the bootstrap method. From the bootstrap method we get the 95% confidence interval for the proportion $CPV(9)/CPV(8)$ is $[1.017; 1.020]$ which means that increasing p from 8 to 9 we will get only 2% more of the variance explained. This is sufficient argument to stop at $p = 8$. Adding more FPC's dose not improve significantly total variance explained.

### 3.3. Application to informative frequency bands

In our experiment, we set the threshold $L$ defined in (5) as 80%. Therefore, for our vibration data coming from the gearbox we will be looking for two sets: $A_{spectr}$ and $A_{FPCA}$ such that $\frac{IE}{E}(A_{spectr}) \geq 80\%$ and $\frac{IE}{E}(A_{FPCA}) \geq 80\%$. We will see how close those two sets are on real applications using vibration signals.

Searching for $A_{spectr}$ that satisfies (5) may be quite time consuming as we have to consider all possible subsets of frequencies from the set $[0, \Lambda]$. For example, for the vibration gearbox data one would have to deal with $2^{513}$ combinations! As described in the previous Section, we start the search by identifying the frequency $f_1$ that maximizes $\sum_{t=0}^{T} |STFT(t,f)|^2$, then select the second in the order of the energy contribution and so on.

The analysis based on $CPV(p)$ presented in the previous subsection has shown that $p = 8$ first functional principal components reproduce as much as 81.88% of the total variability in the data. Therefore, the FPCA method induced by the spectrogram of gearbox signal creates the eight-dimensional vector of scores, taken from the functional expansion. This means reducing the dimensionality of our problem

from 1265 windows with 513 frequencies in each to $8 \times 513$ considering spectrogram data as a functions of frequencies. This means, that finding $A_{FPCA}$ is much faster than $A_{spectr}$, especially for large data sets.

Below, we provide a listing of all frequencies pertaining to $A_{spectr}$ and $A_{FPCA}$.

$$\begin{aligned} A_{spectr} =\{&447.13, 455.11, 439.14, 463.1, 431.16, 471.08, 423.17, 479.06, 415.19, \\ &542.94, 550.92, 534.96, 487.05, 526.97, 558.91, 518.99, 566.89, 407.2, \\ &495.03, 511, 574.88, 503.02, 335.35, 343.33, 327.36, 399.22, 351.31, \\ &319.38, 359.3, 391.24\}. \end{aligned}$$

$A_{spectr}$ has 30 elements.

$$\begin{aligned} A_{FPCA} =\{&447.13, 455.11, 439.14, 463.1, 431.16, 471.08, 423.17, 479.06, 415.19, \\ &542.94, 534.96, 550.92, 487.05, 526.97, 558.91, 407.2, 518.99, 566.89\}. \end{aligned}$$

$A_{FPCA}$ has 18 elements.

Note that the first 10 frequencies coincide, up to the second decimal point (in Hz). If we consider frequencies from the intersection $B = A_{spectr} \cap A_{FPCA}$ , then the percentage of signal energy describing IFB is already at the level of 74,03%, not very far from the threshold of 80%.

We have applied the above analysis to the data set generated by a wheel bearing and described in Cioch et al. (2013). The sets $A_{spectr}$ and $A_{FPCA}$ are shown below for these data

$$\begin{aligned} A_{spectr} = \{&1272.51, 1291.23, 1253.8, 1309.94, 3106.43, 3125.15, 3087.72, 1235.09, 1328.65 \\ &3143.86, 3069.01, 1347.37, 3162.57, 3050.29, 1216.37, 1366.08, 3031.58, 3181.29 \\ &1197.66, 1384.8, \} \end{aligned}$$

$$\begin{aligned} A_{FPCA} = \{&1272.51, 1291.23, 1253.8, 1309.94, 3106.43, 1235.09, 3125.15, 1328.65, 3087.72 \\ &3143.86, 3069.01, 1347.373162.57, 1216.37, 3050.29, 1366.08, 3181.29, 3031.58 \\ &1384.8, 1197.66\}. \end{aligned}$$

A careful examination of the above listings for both data sets shows that both sets contain the same frequencies. They are shown in the order of their importance in energy explained in Section 2. Therefore, the only change we have using the

FPCA is the change of the order of the energy importance of the frequency. Observe, however, that such a change is not dramatic. FPCA preserves the order of the first five frequencies and then makes only small changes, never bigger than two places in the order of energy.

## 4. Conclusions

Our article is devoted to introducing the functional data approach to analyse big data generated by signals available for structural health monitoring. We show that applying the FPCA - the functional principal component approach - we can reduce the dimensionality of the data from several millions to several thousands. Using such an approach we show the importance of the eigenfunctions and eigenvalues calculated for functions generated by observing the signal. It turns out that the popular coefficient - the cumulative percentage of the variance explained (CPV) exceeds 80 per cent for the initial few functional components. We show that this approach applied to the signal generated by the excavating machine can be helpful in identifying informative frequency bands. Moreover, applying the bootstrap approach we can show that the CPV has a relatively small dispersion, which proves the numerical stability of our results.

## Acknowledgement

The Authors would like to express their gratitude to Professor Radoslaw Zimroz from Wrocław for providing access to data regarding gearbox of the excavating machine.

## REFERENCES

BOSQ, D., (2000). Linear Processes in Function Spaces, Springer Verlag.

CHU, F., FENG, Z., LIANG, M., (2013). Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples, Mechanical Systems and Signal Processing, Vol. 38, No. 1, pp. 165–205.

CIOCH, W., KNAPIK, O., LEŚKOW, J., (2013). Finding a frequency signature for a cyclostationary signal with applications to wheel bearing diagnostics, Mechanical Systems and Signal Processing, Vol. 38, pp. 55–64.

GRYLLIAS, K., ANDRE, H., LECLERE, Q., ANTONI, J., (2017). Condition monitoring of rotating machinery under varying operating conditions based

on Cyclo- Non-Stationary Indicators and a multi-order probabilistic approach for instantaneous angular speed tracking, IFAC papers online, Vol. 50-1, pp. 4708–4712.

GUO, X, CHEN, L., SHEN, CH., (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement, Vol. 93, pp. 490–502.

HORVÁTH, L., KOKOSZKA, P., (2012). Inference for Functional Data with Applications, Springer-Verlag, New York etc.

JIA, F., LEI, Y., LIN, J., ZHOU, X., LU, N., (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mechanical Systems and Signal Processing, Vol. 72-73, pp. 303–315.

KHADERSAB, A., SHIVAKUMAR, DR. S., (2018). Vibration Analysis Techniques for Rotating Machinery and its effect on Bearing Faults, Procedia Manufacturing, Vol. 20, pp. 247–252.

KRUCZEK, P., WODECKI, J., WYŁOMAŃSKA, A., (2017). Novel method of informative frequency band selection for vibration signal using nonnegative matrix factorization of short-time fourier transform, IEEE 11th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED).

LIU, R., YANG, B., ZIO, E., CHEN, X., (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review, Mechanical Systems and Signal Processing, Vol. 108, pp. 33–47.

MAS, A., (2002). Weak convergence for the covariance operators of a Hilbertian linear process, Stochastic Processes and their Applications, 99, pp. 117–135.

OBUCHOWSKI, J. WYŁOMAŃSKA, A., ZIMROZ, R., (2014). Selection of informative frequency band in local damage detection in rotating machinery, Mechanical Systems and Signal Processing, 48, pp. 138–152.

RANDALL, B., ANTONI, J., (2011). Rolling element bearing diagnostics - A tutorial, Mechanical Systems and Signal Processing, Vol. 25, pp. 485–520.

RAMSAY, J. O., SILVERMAN, B. W., (2002). Applied functional data analysis, Springer-Verlag.

RAMSAY, J. O., SILVERMAN, B. W., (2005). Functional data analysis, Springer-Verlag.

SPIRIDONAKOS, M. D., FASSOIS, S. D., (2014). Non-stationary random vibration modelling and analysis via functional time-dependent ARMA (FS-TARMA) models - A critical survey, Mechanical Systems and Signal Processing, Vol. 47, pp. 175–224.

STASZEWSKI, W. J., WALLACE, D. M., (2014). Wavelet-based Frequency Response Function for time-variant systems - An exploratory study, Mechanical Systems and Signal Processing, Vol. 47, pp. 35–49.

YANG, Y., NAGARAJAIAH, S., (2014). Blind identification of damage in time-varying systems using independent component analysis with wavelet transform, Mechanical Systems and Signal Processing, Vol. 47, pp. 3–20.