Feature Extraction of Epileptic EEG in Spectral Domain via Functional Data Analysis

Shengkun Xie[1] and Anna Lawniczak[2]
[1]*Ted Rogers School of Management, Ryerson University, Toronto, Canada*
[2]*Department of Mathematics and Statistics, University of Guelph, Guelph, Canada*

Keywords: Functional Data Analysis, Power Spectrum, Functional Principal Component Analysis, EEG, Epilepsy Diagnosis.

Abstract: Functional data analysis is a natural tool for functional data to discover functional patterns. It is also often used to investigate the functional variation of random signals. In this work, we propose a novel approach by analyzing EEG signals in the spectral domain using functional data analysis techniques including functional descriptive statistics, functional probes, and functional principal component analysis. By first transforming EEG signals into their power spectra, the functionality of random signals is greatly enhanced. Because of this improvement, the application of functional data analysis becomes meaningful in feature extraction of random signals. Our study also illustrates a great potential of using functional PCA as a feature extractor for EEG signals in epilepsy diagnosis.

1 INTRODUCTION

Feature extraction of high dimensional data has been an important research area in machine learning (Bouveyron et al., 2007; Kriegel et al., 2009; Yu and Liu, 2003; Jimenez and Landgrebe, 1998). It aims at obtaining a set of key features, so that, the complexity of data classification can be greatly reduced. Ideally, in feature extraction, one looks for a highly separable feature vector as an input for the classification problem. However, the degree of inseparability significantly affects the choice of a classification method. If the extracted features are either linearly or non-linearly separable, there is no extra effort needed to select a suitable classification method. In biomedical signal classification (such as EEG or ECG classification), due to its high dimensional nature, feature extraction of the given signals is often the most important step to meet the success of classification (Alickovic et al., 2018; Truong et al., 2017; Fergus et al., 2016). After the feature vector is obtained, a classification method such as linear discriminate analysis (LDA), k-nearest neighbor (KNN), or support vector machines (SVM) is then applied to determine a group membership.

In many real-world applications of biomedical signal classification (Qazi et al., 2016; Li et al., 2005; Phinyomark et al., 2012; Gandhi et al., 2011; Subasi and Gursoy, 2010), a low dimensional and linear or non-linearly separable feature vector is highly desirable for both, the ease of data visualization in medical devices and a possibility of using a simple classification method, such as a linear classifier or the k-nearest neighbor method. To meet this goal, there is a lot of current research focusing on feature extraction of signals in the time domain using sparse representation of signals (Zhang et al., 2015). This type of research aims at extracting a low dimensional feature vector through sparse decomposition of signals and improvement of the linear or non-linear separability by selecting its most discriminative features. Often the first goal is easy to achieve by enforcing the sparsity on signal approximation, but it is more difficult to make a set of good features, which are linearly or non-linearly separable. Among many published research works, the time-frequency decomposition technique is the most popular within this type. It decomposes the signal in terms of time and frequency domain components. By doing time-frequency domain decomposition, the separability of signals is greatly improved. The classification is then based on extracted features in time and frequency domains, which are the coefficients of the selected time-frequency basis functions (Nyan et al., 2006; Garcia et al., 2003).

Functional data analysis (FDA) is a natural tool for studying functional data, such as images, temper-

Xie, S. and Lawniczak, A.
Feature Extraction of Epileptic EEG in Spectral Domain via Functional Data Analysis.
DOI: 10.5220/0007353301180127
In *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2019)*, pages 118-127
ISBN: 978-989-758-351-3

ature data, or growth curves (Ramsay, 2005; Wang et al., 2016; Ramsay and Silverman, 2007). This is due to the existence of stable, and highly observable deterministic patterns contained in the data, as a part of the key signal features. When data is functional in nature and discriminative, a set of coefficients of basis functions can be obtained by signal representation using common basis functions. From statistical point of view, this belongs to a functional regression problem, and it is usually addressed by an ordinary least square method or a regularization based least square method. The classification of signals can be done by classifying the extracted coefficients obtained from signal representation using basis functions. On the other hand, FDA allows us to investigate the functional variation at various levels. This creates another layer of feature extraction by further looking at the decomposition of data variation. When random signals are not functional in nature, the application of FDA may not be successful. In theory, a random signal can be represented in terms of a linear combination of infinite basis functions. This essentially allows applicability of using FDA for random signals. However, for decomposition of random signals, the coefficients of basis functions, from signal to signal, are often highly volatile, due to the fact that the signals are lacking functional nature. The classification of extracted features may be difficult due to high variability, high dimensionality and lack of separability of the features. This may call for a novel approach in order to improve the functionality of random signals in a given domain.

In this work, we propose a novel approach and illustrate it by analyzing EEG signals in the spectral domain using functional data analysis techniques including functional descriptive statistics, functional probes and functional principal component analysis. We first transform EEG signals into power spectra to improve the functionality of EEG signals. We then use functional PCA to extract signal features. We also demonstrate the application of the proposed method for epilepsy diagnosis (Lima et al., 2009; Liang et al., 2010). The significance of this work is the novelty of the proposed method and the achieved high separability of the considered data, as well as the high applicability of the method to other types of signals such as financial time series and long-term observational economical data for classification or pattern recognition problems. Unlike the regular use of functional principal component analysis, which aims at either estimating the functionality or capturing functional data variation, our work emphasizes the feature extraction of signals. This paper is organized as follows. In Section 2, we discuss the proposed methods including spectral analysis and functional data analysis. In Section 3, the analysis of publicly available EEG data and summary of main results are presented. Finally, we conclude our findings and provide further remarks in Section 4.

2 METHODS

Statistical measures, models or distributions are often used for pattern recognition of high dimensional data or signals. Most of the real-life signals are stochastic and the pattern of how they are related is measured by sample auto-covariance matrix or sample auto-correlation matrix in the time domain. The auto-covariance matrix is then further decomposed using various matrix decomposition methods to reduce the dimension of feature space. For epileptic EEG signals, our study shows that there is a better functional relationship between the periodograms and the frequency values than the pattern in the time domain. This motivates us to study functional data analysis in a spectral domain for clustering EEG signals, instead of analyzing signals in the original time domain. We will first discuss the transformation of a signal to the spectral domain.

2.1 Spectral Analysis

For a given signal X_t of length n, sampled at discrete times, we define the discrete Fourier transform (DFT) to be

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} X_t e^{-2\pi i \omega_j t} \tag{1}$$

for $j = 0, 1, \ldots, n-1$, where the frequency $\omega_j = j/n$. Transforming the signal by discrete Fourier transform allows to obtain a concentration of signal powers using a small set of more dominant frequencies. This means that one is able to focus on a selected ω_j and its transformed values $d(\omega_j)$ only.

By applying the inverse DFT to $d(\omega_j)$, the signal X_t can be exactly recovered as follows:

$$X_t = n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j) e^{2\pi i \omega_j t}. \tag{2}$$

The periodogram for each frequency ω_j is defined as

$$\begin{aligned} I(\omega_j) &= |\, d(\omega_j)\,|^2 \\ &= \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} (X_t - \bar{X})(X_s - \bar{X}) e^{-2\pi i \omega_j (t-s)} \\ &= \frac{1}{n} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X})(X_s - \bar{X}) e^{-2\pi i \omega_j h} \\ &= \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-2\pi i \omega_j h}, \end{aligned}$$

where $\hat{\gamma}(h)$ is the auto-covariance function of time lag h. In the above, we used the fact that

$$\sum_{t=1}^{n} e^{-2\pi i\omega_j t} = 0. \tag{3}$$

Notice that the periodogram is just the Fourier transform of the auto-covariance function, which captures the quadratic covariation of signal in the spectral domain.

Furthermore, we can express the periodogram as a sum of squared sine and cosine transforms of a signal as follows:

$$|\, d(\omega_j)\,|^2 = d_c^2(\omega_j) + d_s^2(\omega_j) \tag{4}$$

where $d_c(\omega_j)$ is the cosine transform and $d_s(\omega_j)$ is the sine transform of the signal. These transforms are defined, respectively, as follows:

$$d_c(\omega_j) = n^{-1/2} \sum_{t=1}^{n} X_t \cos(2\pi\omega_j t) \tag{5}$$

$$d_s(\omega_j) = n^{-1/2} \sum_{t=1}^{n} X_t \sin(2\pi\omega_j t), \tag{6}$$

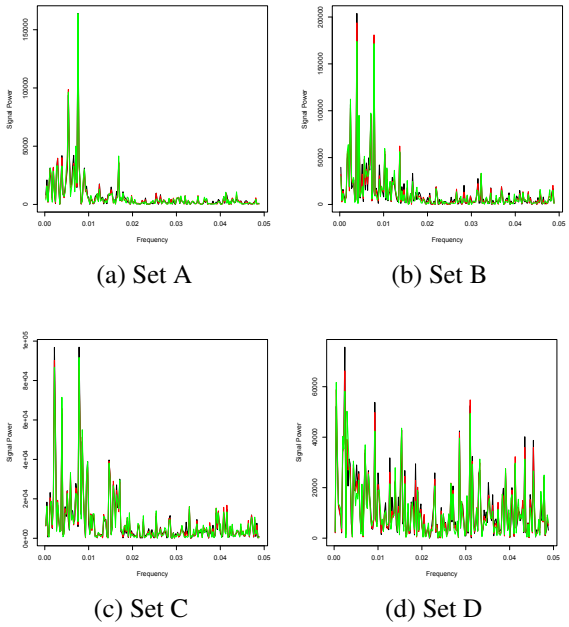
where $j = 0, 1, 2, \ldots, n-1$. $I(\omega_j)$ is also called a power spectrum of a signal X_t. Because of its definition, a periodogram captures the distribution of variation of a signal in the spectral domain. The larger a value of the periodogram the more dominant is its corresponding frequency. Thus, the dominant values determine the signal power spectra. Often these more dominant frequencies correspond to smaller frequency values, which implies that local patterns are more significant than the global one, after the signal is transformed into the spectral domain. Because of this, we only focus on the analysis of the power spectrum in a sub-interval, i.e. we analyze only the first 200 frequency values for the given signals.

To illustrate the above points, we present a set of results of the power spectra for selected different types of EEG signals in Figure 1 (the data set will be discussed later). One can clearly see that the power spectra of the first 200 frequency values behave similarly within each set of data, but their patterns look differently over different frequency values.

2.2 Functional Data Analysis

Often a signal is sampled at discrete times. However, functional data analysis allows us to model the power spectrum of a signal with discrete observations by a linear combination of a set of continuous basis functions. Mathematically, for the ith signal in a given data set, we can expand the power spectrum $I_i(\omega)$ by

$$I_i(\omega) = \sum_{k=1}^{K} \alpha_{ik}\phi_k(\omega), \tag{7}$$



(a) Set A (b) Set B

(c) Set C (d) Set D

Figure 1: Sample plots of the power spectra for data sets A (Normal: Eyes Closed), B (Normal: Eyes Open), C (Non-epileptogenic zone) and D (Epileptogenic zone), respectively. Each plot contains three sample power spectra of the first 200 frequency values.

where ω is the frequency value, α_{ik} is the coefficient of the kth basis function and K is the total number of basis functions. In feature extraction, since our objective is not to fully represent a power spectrum using a set of functional basis, often a small finite number K is chosen. That is, we approximate the power spectrum by a linear combination of a small number of K basis functions. Notice that, within the discussion of this section, we do not separate the mean function from the representation of a signal, but in the later discussion we will separate the mean function from the signal expansion because we will study the functional variation of the signal power spectra.

For a sample of N signals, the matrix notation of power spectrum becomes

$$\mathbf{I}(\omega) = \mathbf{C}\Phi(\omega) \tag{8}$$

where $\mathbf{I}(\omega)$ is a column vector of length N, $\Phi(\omega)$ is a column vector of length K containing the basis functions, and $\mathbf{C}$ is the coefficient matrix of size $N \times K$. Notice that the basis function $\phi(\omega)$ can be different for various groups of signals. However, because we will consider the signals which share many commonalities in the spectral domain, it is more reasonable to use the same basis functions for all the considered groups of signals. This will allow us to extract signal features within the same feature space. In functional data analysis, there are two types of basis func-

tions: periodical basis functions and non-periodical basis functions. In this work, we select non-periodical basis functions as we don't expect that power spectra to be periodic. Within non-periodical basis functions, B-spline basis function is the most popular one, therefore we choose this type of basis function for our investigation (De Boor et al., 1978; Unser et al., 1993). Of course, we cannot rule out the possibility of the effect of selected basis functions, so we will investigate this effect in this work.

2.2.1 Functional Descriptive Statistics

Summary statistics are often the key features, which are useful for signal classification. Since we have the functional representation of data, the basic summary statistics can be expressed in functional form as a function of frequency ω.

For a given set of sample of length N, the functional mean power spectrum is given as follows:

$$\begin{aligned}\bar{I}(\omega) &= N^{-1}\sum_{i=1}^{N} I_i(\omega) = N^{-1}\sum_{i=1}^{N}\sum_{k=1}^{K}\alpha_{ik}\phi_k(\omega)\\ &= \sum_{k=1}^{K}\bar{\alpha}_k\phi_k(\omega), \qquad (9)\end{aligned}$$

where $\bar{\alpha}_k$ is the average of the coefficients at a given frequency of the N signals. When $N\to\infty$, $\bar{\alpha}_k\to\alpha_k$, where α_k is the true coefficient of $\phi_k(\omega)$. In this case, we can denote the mean power spectrum by

$$\mu(\omega) = \sum_{k=1}^{K}\alpha_k\phi_k(\omega). \qquad (10)$$

Similarly to the sample variance, the functional variance, calculated based on N sample power spectra, is given by

$$\begin{aligned}S^2_{I_{i(\omega)}} &= \frac{1}{N-1}\sum_{i=1}^{N}\left(I_i(\omega)-\bar{I}(\omega)\right)^2\\ &= \frac{1}{N-1}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}\alpha_{ik}\phi_k(\omega)-\sum_{k=1}^{K}\bar{\alpha}_k\phi_k(\omega)\right)^2\\ &= \frac{1}{N-1}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}(\alpha_{ik}-\bar{\alpha}_k)\phi_k(\omega)\right)^2. \qquad (11)\end{aligned}$$

For a given sample of N signals, the functional variance-covariance at two different frequency values ω_1 and ω_2 can be estimated by

$$\begin{aligned}v(\omega_1,\omega_2) &= \frac{1}{N-1}\sum_{i=1}^{N}\left(I_i(\omega_1)-\bar{I}(\omega_1)\right)\\ &\quad\left(I_i(\omega_2)-\bar{I}(\omega_2)\right). \qquad (12)\end{aligned}$$

2.2.2 Functional Probes

Purely descriptive statistics such as functional mean, functional variance, or functional covariance, allow us to see the functional central tendency and the functional variation pattern of signal power spectra. However, they are high dimensional statistics. In signal classification, the dimensions of these functional descriptive statistics must be further reduced. Here, we consider an application of functional probes. A probe ρ_ξ is a measure allowing us to see specific variation by defining a functional weight $\xi(\omega)$, and it is defined as

$$\rho_\xi = \int \xi(\omega)I(\omega)d\omega. \qquad (13)$$

This is an inner product of functions $\xi(\omega)$ and $I(\omega)$. The $\xi(\omega)$ has been structured so that we can extract specific features or patterns of the variation in power spectrum $I(\omega)$. In this work, we choose functional mean and functional standard deviation of the power spectrum from a given group of signals as weight functions. The probe values for the ith signal power spectrum using functional mean and functional standard deviation of the jth group power spectrum becomes

$$\begin{aligned}\rho_{\bar{I}_{ij}} &= \int \bar{I}^{(j)}(\omega)I_i(\omega)d\omega\\ &= \sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\alpha_{ik_1}\bar{\alpha}^{(j)}_{k_2}\int\phi_{k_1}(\omega)\phi_{k_2}(\omega)d\omega,\\ \rho_{S_{ij}} &= \int S^{(j)}_{I_{i(\omega)}}I_i(\omega)d\omega. \qquad (14)\end{aligned}$$

The functional probe values capture the similarity between the weight function and the ith power spectrum of a signal. When the basis functions are orthonormal (i.e., $\int\phi_{k_1}(\omega)\phi_{k_2}(\omega)d\omega = 0$, for $k_1\neq k_2$, and $\int\phi^2_k(\omega)d\omega = 1$, for $k = 1,2,\ldots,K$), the probe value using a functional mean becomes

$$\rho_{\bar{I}_{ij}} = \sum_{k=1}^{K}\alpha_{ik}\bar{\alpha}^{(j)}_k, \qquad (15)$$

which can be interpreted as a similarity measure between two different groups of signals in the spectral domain. However, the closed form does not exist for probe value using functional standard deviation.

So far, we have discussed functional probe values based on the power spectrum $I(\omega)$. If we replace $I(\omega)$ by $v(\omega_1,\omega_2)$, i.e. the variance-covariance function, the functional probe value becomes

$$\int\xi(\omega_2)v(\omega_1,\omega_2)d\omega_2. \qquad (16)$$

This is exactly the left hand side of the eigen-equation for solving eigenvalues and eigenvectors in functional principal component analysis, which will be discussed in the next section.

2.2.3 Functional Principal Component Analysis

In multivariate statistics, principal component analysis (PCA) of p-variate random vector $X = (X_1, X_2, \ldots, X_p)$ looks for a set of weight values, ξ_j=$(\xi_{1j}, \xi_{2j}, \ldots, \xi_{pj})$, so that, at the jth step, the linear combination of variables X_i has the greatest variance. That is $Var(\sum_{i=1}^{p} \xi_{ij} X_i)$, or in the matrix notation, $Var(\xi_j^\top X)$ is maximized. For $j = 1, 2, \ldots, p$, this process is repeated by replacing each X_i by the value obtained after subtracting the previous principal component, subject to $\sum_{i=1}^{p} \xi_{ij}^2 = 1$ and $\sum_{i=1}^{p} \xi_{ij}\xi_{il} = 0$ for $j < l$, and $1 \leq l, j \leq p$. The actual implementation of this procedure can be done by a singular value decomposition (SVD) of the data matrix $\mathbf{X}$, which contains N realizations of X. In this work, data matrix $\mathbf{X}$ becomes $\mathbf{I}(\omega)$, which is also an $N \times p$ data matrix, where p is the total number of frequency values being considered. Formally, the SVD of $\mathbf{I}(\omega)$ is a factorization of the form $\mathbf{U\Sigma V}$, where $\mathbf{U}$ is a $N \times N$ unitary matrix, Σ is a $N \times p$ matrix consisting of eigenvalues of $\mathbf{I}(\omega)$, and $\mathbf{V}$ is a $p \times p$ unitary matrix. The columns of $\mathbf{U}$ and the columns of $\mathbf{V}$ are called the left eigenvector and the right eigenvector of $\mathbf{I}(\omega)$, respectively. Also, each column of $\mathbf{V}$ is just the weight vector ξ_j. The feature extraction of data matrix $\mathbf{I}(\omega)$ becomes the computation of $\mathbf{I}(\omega)\xi_j^\top$, for $j = 1, 2, \ldots, p$. For example, the first principal component scores set is $\mathbf{I}(\omega)\xi_1^\top$, and the second principal component scores set is $\mathbf{I}(\omega)\xi_2^\top$.

Notice that, the functional probes discussed above aim at capturing the variation of data associated with the weight function. If we carefully select the functional weight $\xi(\omega)$, so that the variance of functional probe values in (13) is maximized, subject to the constraint that $\int \xi_l(\omega)\xi_j(\omega)d\omega = 0$ for $l \neq j$, and $\int \xi^2(\omega)d\omega = 1$, then this becomes the functional PCA. In this case, functional probe values are the principal component scores and the weight function becomes functional principal component loading. From the discussion above, we can see the connection between the multivariate PCA and functional PCA. If we focus on only the discrete values of ω, then functional PCA becomes multivariate PCA. However, the functional PCA allows us to explore the functional variation of different principal components.

Suppose that the power spectrum of a given signal can be expanded using K basis functions, which is given as follows:

$$I_i(\omega) = \mu(\omega) + \sum_{k=1}^{K} \beta_{ik}\phi_k(\omega), \tag{17}$$

where $\mu(\omega)$ is the functional mean of power spectrum. We may then express this in a matrix notation

$$\mathbf{I} - \mu = \mathbf{C}\phi, \tag{18}$$

where $\mathbf{C}$ is the $N \times K$ coefficient matrix, and ϕ=$(\phi_1, \phi_2, \ldots, \phi_K)^\top$. Now, we consider how to obtain the function principal components and their scores. First let us denote the variance-covariance function by $v(\omega_1, \omega_2)$. This function is defined in (12). In matrix form the variance-covariance function is

$$v(\omega_1, \omega_2) = \frac{1}{N-1}\phi^\top(\omega_1)\mathbf{C}^\top\mathbf{C}\phi(\omega_2). \tag{19}$$

Next, to find the principal component weight functions, we have to solve the following eigen-equation for the appropriate eigenvalue λ

$$\int v(\omega_1, \omega_2)\xi(\omega_2)d\omega_2 = \lambda\xi(\omega_1). \tag{20}$$

Suppose that the eigen-function $\xi(\omega)$ has an expansion

$$\xi(\omega) = \sum_{k=1}^{K} b_k\phi_k(\omega), \tag{21}$$

or in the matrix notation

$$\xi(\omega) = \phi^\top(\omega)\mathbf{b}, \tag{22}$$

where $\mathbf{b}$=$(b_1, b_2, \ldots, b_K)$. This yields

$$\begin{aligned}\int v(\omega_1, \omega_2)\xi(\omega_2)d\omega_2 &= \frac{1}{N-1}\int \phi^\top(\omega_1)\mathbf{C}^\top\mathbf{C} \\ &\quad \phi(\omega_2)\phi^\top(\omega_2)\mathbf{b}d\omega_2. \\ &= \frac{1}{N-1}\phi^\top(\omega_1)\mathbf{C}^\top\mathbf{C}\Phi\mathbf{b},\end{aligned}$$

where Φ=$\int \phi(\omega)\phi^\top(\omega)d\omega$ is a $K \times K$ matrix. Thus, the eigen-equation (20) becomes

$$\frac{1}{N-1}\phi^\top(\omega)\mathbf{C}^\top\mathbf{C}\Phi\mathbf{b} = \lambda\phi^\top(\omega)\mathbf{b}. \tag{23}$$

Since equation (23) must hold for all ω, this implies the following matrix equation

$$\frac{1}{N-1}\mathbf{C}^\top\mathbf{C}\Phi\mathbf{b} = \lambda\mathbf{b}. \tag{24}$$

To obtain the required principal components, we define $\mathbf{u} = \Phi^{1/2}\mathbf{b}$, thus the equation (24) becomes

$$\frac{1}{N-1}\Phi^{1/2}\mathbf{C}^\top\mathbf{C}\Phi^{1/2}\mathbf{u} = \lambda\mathbf{u}. \tag{25}$$

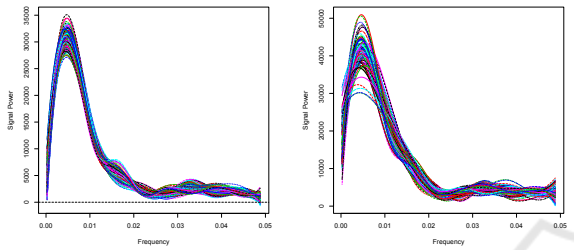
By solving the symmetric eigenvalue problem in (25) for $\mathbf{u}$, and then computing $\mathbf{b} = \Phi^{-1/2}\mathbf{u}$ one gets the eigen-function $\xi(\omega)$, which is given by

$$\xi(\omega) = \phi^\top(\omega)\Phi^{-1/2}\mathbf{u}. \tag{26}$$

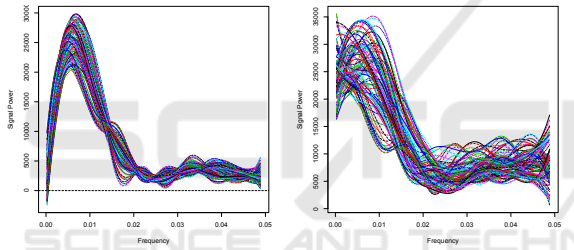
If $\phi_k(\omega)$ are orthogonal, then Φ becomes the $K \times K$ identity matrix. Thus, the eigen-analysis of the functional PCA problem in (24) reduces to

$$\frac{1}{N-1}\mathbf{C}^\top\mathbf{C}\mathbf{b} = \lambda\mathbf{b}, \tag{27}$$

which is the multivariate PCA that replaces variance-covariance matrix by the coefficient matrix $\mathbf{C}$ obtained from the function approximation of power spectrum. From the discussion above, we notice that the multivariate PCA conducts eigen-analysis for a $p \times p$ covariance matix. With the function approximation using K basis functions, the eigen-analysis of functional PCA is applied to a $K \times K$ coefficient matrix, which depends on the value of K. In the case of using sparse approximation, which gives a small value of K, the problem is more efficient in terms of computational complexity.



(a) Set A (b) Set B



(c) Set C (d) Set D

Figure 2: The plots of functional power spectrum for data sets A (Normal: Eyes Closed), B (Normal: Eyes Open), C (Non-epileptogenic zone) and D (Epileptogenic zone) respectively. 10 B-splines basis functions are used to smooth sample power spectra.

2.2.4 Feature Extraction by Functional Principal Component Analysis

After the eigen-function $\xi(\omega)$ is obtained, we can extract the principal component scores, denoted by P_j, for the given power spectrum $\mathbf{I}(\omega)$ by the following formula

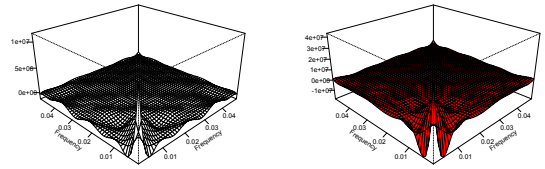
$$P_j = \int \mathbf{I}(\omega)\xi_j(\omega)d\omega, \ \ j = 1,\ldots,K. \tag{28}$$

Subsitituting (18) and (26) into the equation above, we get

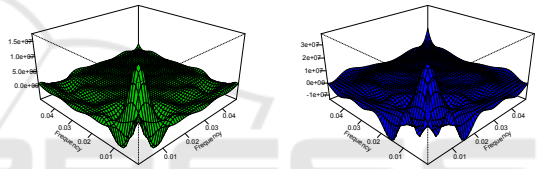
$$\begin{aligned} P_j &= \int (\mu(\omega) + \mathbf{C}\phi(\omega))\phi^\top(\omega)\Phi^{-1/2}\mathbf{u}_j d\omega. \quad (29) \\ &= \int \mu\phi^\top\Phi^{-1/2}\mathbf{u}_j + \int \mathbf{C}\phi\phi^\top\Phi^{-1/2}\mathbf{u}_j \\ &= \bar{\mu}\Phi^{-1/2}\mathbf{u}_j + \mathbf{C}\Phi^{-1/2}\mathbf{u}_j, \quad (30) \end{aligned}$$

where $\bar{\mu} = \int \mu(\omega)\phi^\top(\omega)d\omega$. Thus, P_1 to the first principal component score vector of the N signal power spectra, and P_2 is the second principal component score vector, and so on.

Notice that, to fully represent a given power spetrum $I_j(\omega)$, the number of basis functions K may approach to infinity. However, for the purpose of feature extraction, we aim for a low dimensional feature subspace, which may require a choice of small K value. We then use a cubic B-spline basis, which gives the order number to be 4, and we select the number of basis functions K to be 10. This selection leads to a good approximation of the power spectra.



(a) Set A (b) Set B

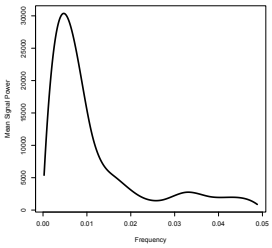


(c) Set C (d) Set D

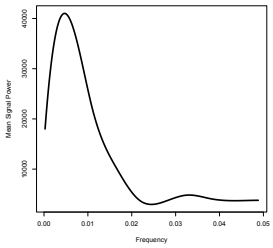
Figure 3: The plots of functional covariance of power spectra for data sets A (Normal: Eyes Closed), B (Normal: Eyes Open), C (Non-epileptogenic zone) and D (Epileptogenic zone), respectively.

3 RESULTS

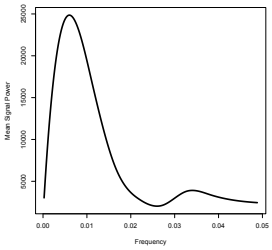
In order to demonstrate the application of the proposed method to epilepsy diagnosis, we use a set of EEG signals coming from healthy volunteers and from patients during seizure-free intervals. This database is from the University of Bonn, Germany (http://epileptologie-bonn.de/cms/front_content.php?idcat=193). There are four different sets of EEG data, denoted, respectively by A, B, C, and D. Data in sets A and B are normal signals with eyes closed and open, respectively. Data in sets C and D are epileptic signals coming from patients suffering from epilepsy. Signals in the set C are collected from the patients' non-epileptogenic zone, and signals in the set D are from the patients' epileptogenic zone. Each dataset contains 100 single channel scalp EEG segments of a 23.6 second duration. To achieve this the EEG signals, were sampled at 173.61 Hz (i.e., T=4096).



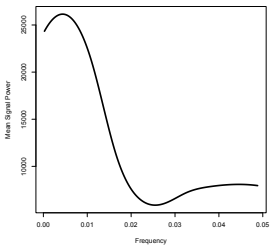
(a) Set A



(b) Set B

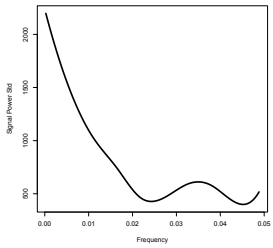


(c) Set C

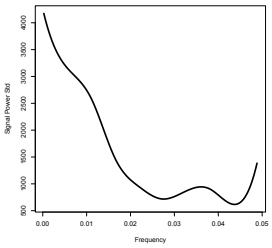


(d) Set D

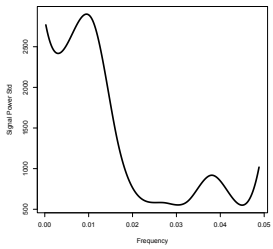
Figure 4: The plots of functional mean power spectra for data sets A (Normal: Eyes Closed), B (Normal: Eyes Open), C (Non-epileptogenic zone) and D (Epileptogenic zone), respectively.



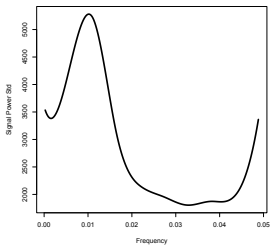
(a) Set A



(b) Set B



(c) Set C

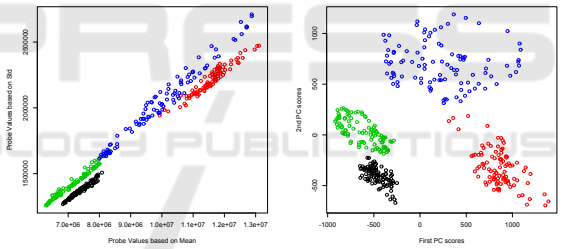


(d) Set D

Figure 5: The plots of the functional standard deviations of power spectra for data sets A (Normal: Eyes Closed), B (Normal: Eyes Open), C (Non-epileptogenic zone) and D (Epileptogenic zone), respectively.

From the results displayed in Figure 2, one can see that the smoothed signal power spectra behave similarly for both signals from healthy people (sets A and B) and signals from patients, which were collected from a non-epileptogenic zone (set C). However, there are still some differences that we can see among the graphs. This may suggest that further classification is needed based on these power spectra to recognize the differences hidden in the power spectra. Also, one can see that the power spectra of signals collected from patients' epileptogenic zone (set D) are more volatile and look different from the signals of healthy people. However, they share some commonalities with signals from the set C. The graphical dispay offers some evidence that suitable clustering methods may differentiate these types of signals successfully.

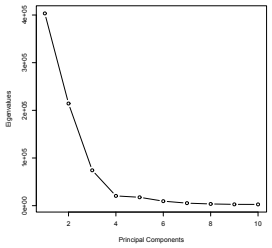
Many clustering methods, including PCA are applied to the variance-covariance data matrix to recognize the differences among different groups of data. Therefore, we further examine the auto-covariance data matrices for each set of signals. The results displayed in Figure 3 show that the auto-covariances are significant only at low-frequency values. This suggests that it may be sufficient to extract features from small signal windows only. Thus, the focus on more dominant powers within the spectral domain may lead to a dimension reduction and this expectation is confirmed by our results displayed in Figure 1.



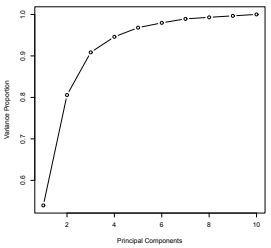
(a) Functional Probes (b) PC scores

Figure 6: The extracted probe values using the functional mean and the functional standard deviation of power spectrum of signals of the Set D as the functional probe for data sets A (Normal: Eyes Closed, Black), B (Normal: Eyes Open, Red), C (Non-epileptogenic zone, Green) and D (Epileptogenic zone, Blue) are displayed in (a). The plot for extracted principal component scores for first two components of power spectra for data sets A (Normal: Eyes Closed, Black), B (Normal: Eyes Open, Red), C (Non-epileptogenic zone, Green) and D (Epileptogenic zone, Blue) are displayed in (b).

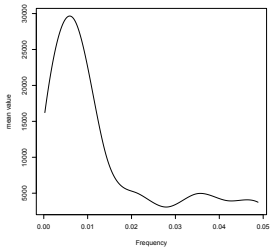
In Figures 4 and 5, the functional mean and the functional standard deviation are reported for each set of signals. In these figures, we observe the strong commonalities and similarities among the respective functional statistics of the sets A, B, and C, but these statistics look different from the ones of the set D. This may imply a different nature of EEG signals among the patients and non-patients (including sig-



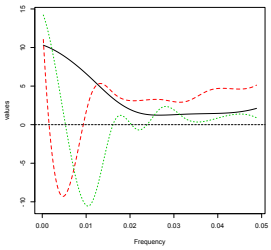
(a) Eigenvalues



(b) Cumulative Variance

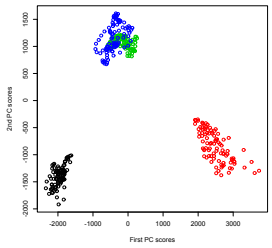


(c) Overall Functional Mean

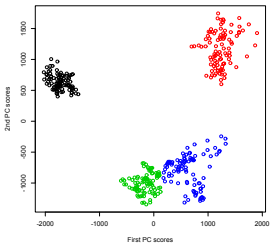


(d) First Three Harmonics

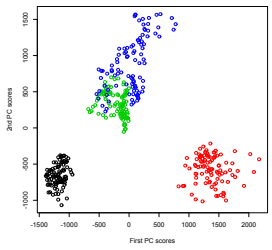
Figure 7: The results of principal components of power spectra including eigenvalues, variance proportion, overall functional mean and the first three eigenfunctions. The first eigenfunction is in black, the second eigenfunction is in red and the third one is in green.



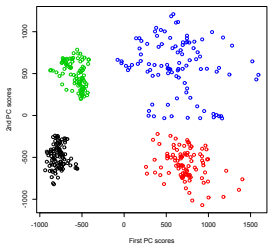
(a) K=200



(b) $K = 100$



(c) $K = 50$



(d) $K = 25$

Figure 8: The Evolution of extracted first two principal component scores of power spectra under different choices of K for data sets A (Normal: Eyes Closed, Black), B (Normal: Eyes Open, Red), C (Non-epileptogenic zone, Green) and D (Epileptogenic zone, Blue).

nals from non-epileptogenic zone). Also, we observe that the functional variation at high frequency is much higher for patitents' data, particularly for the set D. To further reduce the dimensionality of the power spectrum and its functional mean and its functional standard deviation, the functional probe values are calculated based on the inner product of a selected functional mean and a given signal power spectrum. The results using functional standard deviation as a weight function are also obtained. In our study, the best results, in terms of separability of features, are the ones that use the functional mean and the functional standard deviation calculated from the set D. Using these two functional probes, we extract a two-dimensional feature vector from a given signal power spectrum. The Figure 6 (a) clearly display the pattern, which shows a great separability of extracted features (i.e., functional probe values), due to the dimension reduction. We also see that the relationship between these two features seem to be linear for all cases. This implies that the feature variation within each group might not have been maximized, due to the pre-fixed functional probes. When compared to the results obtained by PCA, we confirm that indeed this is the case.

Principal component extractions aim at obtaining a set of scores, so that, further investigation can be done by using them. Our goal is to obtain a set of principal components scores that form clusters for different groups of signals. The obtained principal component scores are reported in Figure 6 (b) under the selection of $K = 10$, which is considered to be the optimal value in terms of feature separability. From the displayed results, we can see that in clustering power spectra of the four sets of signals the functional PCA is more successful than the functional mean and the functional standard deviation as the probes, as can be seen from the previously discussed results. We notice that, the principal components scores of power spectra are completely separable from other groups. Overall, the principal components scores form better into clusters that have clear centroids and more homogenous variations. This is particularly important and useful for model-based classification methods such as Gaussian mixture models. The Figure 7 displays more results of the principal component analysis including eigvenvalues, cumulative variance proportion, functional grand mean and functional principal components. The displayed results show that the first three principal components are dominant in terms of explanation of data variation. The functional grand mean reflects, mainly, the pattern of the power spectra of the sets A, B, and C. These sets have either signals from healthy people or signals from a non-epileptogenic zone of a patient. The main characteristics of power spectra of the set D are reflected in both second and third principal components, while

the first principal component captures the functional grand mean. This helps to explain why by taking the functional mean and the functional standard deviation of power spectrum from set D as functional probes also produce separable functional probe values.

We further investigate the effect of the number of basis functions (i.e., K) on the separability of extracted signal features (i.e., principal component scores of power spectra). The obtained results are displayed in Figure 8. We observe that the proposed method is highly successful in separating the artifacts (i.e., open/closed eyes) as the results did not depend on how the number of basis functions was selected. The feature separability increases with the decrease of K, i.e., the number of basis functions. This may suggest that the sparsity in approximation of the signal power spectra plays an important role in the success of applying functional principal component analysis. When K=200, the extracted features for epileptic signals overlap significantly. This overlapping changes when K decreases, and features start to be fully separable when K is relatively small, for example, around 25. However, the overall separability between healthy and epileptic signals is not affected by the number of basis functions.

4 CONCLUDING REMARKS

Clustering and classification of highly dimensional data are important tasks in pattern recognition and artificial intelligence. To be successful in using machine learning techniques, including clustering and classification, dimension reduction of data is a key approach. In this work, we have conisdered an approach that first transforms signals to the spectral domain and obtains their power spectra. Next, we have applied the functional data analysis techniques to further investigate the charateristics of the signals. We have demonstrated that functional data analysis in spectral domain is useful for understanding the key features of different types of EEG signals. Especially, the extracted features, using functional principal component analysis, can be used for classification of different types of EEG signals. Also we have investigated the effect of sparsity on the performance of separating signal features. The obtained results demonstrate that the proposed method may be useful for an epilepsy diagnosis. Future work will focus on the study of wavelet spectral domain functional PCA and its application to clustering random signals.

ACKNOWLEDGEMENT

The authors acknowledges partial support from NSERC (Natural Sciences and Engineering Research Council of Canada).

REFERENCES

Alickovic, E., Kevric, J., and Subasi, A. (2018). Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomedical Signal Processing and Control*, 39:94–102.

Bouveyron, C., Girard, S., and Schmid, C. (2007). High-dimensional data clustering. *Computational Statistics & Data Analysis*, 52(1):502–519.

De Boor, C., De Boor, C., Mathématicien, E.-U., De Boor, C., and De Boor, C. (1978). *A practical guide to splines*, volume 27. Springer-Verlag New York.

Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K., and Hamdan, H. (2016). A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics*, 12(1):70–89.

Gandhi, T., Panigrahi, B. K., and Anand, S. (2011). A comparative study of wavelet families for eeg signal classification. *Neurocomputing*, 74(17):3051–3057.

Garcia, G. N., Ebrahimi, T., and Vesin, J.-M. (2003). Support vector eeg classification in the fourier and time-frequency correlation domains. In *Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*, pages 591–594. IEEE.

Jimenez, L. O. and Landgrebe, D. A. (1998). Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 28(1):39–54.

Kriegel, H.-P., Kröger, P., and Zimek, A. (2009). Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(1):1.

Li, D., Pedrycz, W., and Pizzi, N. J. (2005). Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification. *IEEE Transactions on biomedical engineering*, 52(6):1132–1139.

Liang, S.-F., Wang, H.-C., and Chang, W.-L. (2010). Combination of eeg complexity and spectral analysis for epilepsy diagnosis and seizure detection. *EURASIP Journal on Advances in Signal Processing*, 2010:62.

Lima, C. A., Coelho, A. L., and Chagas, S. (2009). Automatic eeg signal classification for epilepsy diagnosis with relevance vector machines. *Expert Systems with Applications*, 36(6):10054–10059.

Nyan, M., Tay, F., Seah, K., and Sitoh, Y. (2006). Classification of gait patterns in the time–frequency domain. *Journal of biomechanics*, 39(14):2647–2656.

Phinyomark, A., Phukpattaranont, P., and Limsakul, C. (2012). Feature reduction and selection for emg signal classification. *Expert Systems with Applications*, 39(8):7420–7431.

Qazi, K. I., Lam, H., Xiao, B., Ouyang, G., and Yin, X. (2016). Classification of epilepsy using computational intelligence techniques. *CAAI Transactions on Intelligence Technology*, 1(2):137–149.

Ramsay, J. (2005). Functional data analysis. *Encyclopedia of Statistics in Behavioral Science*.

Ramsay, J. O. and Silverman, B. W. (2007). *Applied functional data analysis: methods and case studies*. Springer.

Subasi, A. and Gursoy, M. I. (2010). Eeg signal classification using pca, ica, lda and support vector machines. *Expert systems with applications*, 37(12):8659–8666.

Truong, N. D., Kuhlmann, L., Bonyadi, M. R., Yang, J., Faulks, A., and Kavehei, O. (2017). Supervised learning in automatic channel selection for epileptic seizure detection. *Expert Systems with Applications*, 86:199–207.

Unser, M., Aldroubi, A., Eden, M., et al. (1993). B-spline signal processing: Part i theory. *IEEE transactions on signal processing*, 41(2):821–833.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295.

Yu, L. and Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 856–863.

Zhang, Z., Xu, Y., Yang, J., Li, X., and Zhang, D. (2015). A survey of sparse representation: algorithms and applications. *IEEE access*, 3:490–530.