



Fourier Spectral Domain Functional Principal Component Analysis of EEG Signals

Shengkun Xie¹(✉) and Anna T. Lawniczak²

¹ Department of Global Management Studies, Ted Rogers School of Management,
Ryerson University, Toronto, ON M5B 2K3, Canada

shengkun.xie@ryerson.ca

² Mathematics and Statistics,

University of Guelph, Guelph, ON N1G 2W1, Canada

alawnicz@uoguelph.ca

Abstract. Functional principal component analysis (FPCA) is a natural tool for investigating the functional pattern of functional data. Clustering functional data using FPCA can be an important area in machine learning for signal processing, in particular, for signals that contain a deterministic component. Also, FPCA can be useful for analysis of random signals if an appropriate transformation is applied. In this work, we propose a novel approach by extracting key features of EEG signals in the Fourier spectral domain using FPCA. By first transforming EEG signals into their Fourier power spectra, the functionality of signals is greatly enhanced. Due to this improvement, the application of FPCA becomes much more meaningful in signal feature extraction. Our study shows a great potential of using spectral domain FPCA as a feature extractor for processing EEG signals in both epilepsy diagnosis and epileptic seizure detection.

Keywords: Fourier power spectrum · Functional principal component analysis EEG · Epilepsy diagnosis

1 Introduction

Feature extraction of high-dimensional data has been an important research area in machine learning [2, 7, 8, 27]. The traditional problem of feature extraction of high-dimensional data focuses on the study of multivariate data with a goal of reducing the dimensionality from the original observation domain to a feature domain. This problem is typically encountered in two machine learning tasks, clustering and classification, depending on whether the data labelling is available or not. Clustering is an unsupervised learning technique that aims at grouping the multivariate data into a selected number of clusters, while classification focuses on predicting the group membership of test data by first learning a mathematical model based on the training data and then using this model to make

a prediction. The performance of clustering or classification is often affected by the dimension of multivariate data. This is why the reduction of data dimension is firstly considered in any given problem. Random signals such as EEG or financial time series are special type of high-dimensional data that often contain no deterministic patterns. Instead, the pattern may appear to be stochastic, which is often time dependent if the random signals are observations over time. Moreover, in real-world problems, most of the signals are non-stationary, which makes the time dependency difficult to measure because of the instability or containing too much uncertainty. Due to the fact of lacking a stable pattern that can be modelled by some mathematical equations, the clustering or classification of random signals directly within the original time domain is challenging. Therefore, feature extraction of random signals is usually the most important step to meet the success of classification or clustering [1, 4, 18, 22, 23], and the development of new methodology for this type of problem becomes highly desirable.

In real-world applications of biomedical signal classification [5, 9, 13, 14, 17], a low-dimensional and linearly or non-linearly separable feature vector is highly desirable for both, the ease of data visualization in medical devices and the possibility of using simple classification methods, such as a linear classifier or the k-nearest neighbour (k-NN) 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 [28]. The idea of sparse representation is to approximate a given signal using a set of basis functions and to obtain a number of coefficients related to the approximation. The obtained coefficients of basis functions are used as signal features. The further study of signal characteristics is then based on these extracted features. However, a set of time domain basis functions used for approximating a given signal may not be a good choice for another signal, thus the extracted features may appear to be extremely volatile. This may lead to a low performance when using these extracted features in applications. To overcome this, the focus of sparse representation of signals has been moved from the time domain to other single or multiple domains. Among many published research works, the time-frequency domain decomposition is the most popular one within this type of approach. It decomposes the signal in terms of time and frequency domain components. By extending the analysis from a single domain to multiple domains, the separability of signal features is often significantly improved and the classification based on extracted features in both time and frequency domains will outperform a single domain approach [6, 12].

For sparse representation of signals, either in time domain or time-frequency domain, the objective is to achieve an explainable or low-dimensional feature vector, so that classification or clustering of signals becomes easier in feature domain [24]. When these signals contain deterministic patterns and the signal to noise ratio is high, they are typically referred to functional data. The functional data can be in spatial domain, temporal domain or both. Some examples of functional data include, but are not limited to images, temperature data, and growth curves [15, 16, 20]. To study functional data, functional data analysis (FDA) is a natural tool, due to its capability of capturing stable, and observable

deterministic patterns contained in the data. In [25], we have used functional data analysis and applied functional summary statistics, functional probes and functional principal components to both epileptic EEG signals without seizure and non-epileptic signals. We have demonstrated that feature extraction through the functional data analysis is able to produce low-dimensional and explainable features so that signals of different types can be discriminated. In this extended work, we will further present the results using the epileptic EEG signals with seizure onset. We will focus on the study of using functional principle component analysis by providing more technical details. In particular, we will provide on discussion of the effect of number of basis functions retained in functional PCA for clustering. Also, we will demonstrate the application of the proposed method for both the epilepsy diagnosis and epileptic seizure detection [10, 11, 21, 26].

This paper is organized as follows. In Sect. 2, we discuss the proposed methods including Fourier power spectra and functional principal component analysis. In Sect. 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 Sect. 4.

2 Methods

2.1 Fourier Power Spectra

To define the Fourier power spectra of signals, we have to first apply the Fourier transformation. For a given signal X_t of length n , sampled at discrete times, the discrete Fourier transform (DFT) is defined as

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

for $j = 0, 1, \dots, n - 1$, and 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. The signal powers that correspond to the selected frequencies provide us a first layer of feature extraction, and due to the focus on the selected frequencies, the feature sparsity is achieved. However, direct use of Fourier transformed values for the given signal is very inconvenient, because they are complex numbers. A signal periodogram is then created to avoid this difficulty. The periodogram for each frequency ω_j is defined as

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

where $\hat{\gamma}(h)$ is the auto-covariance function of time lag h . From the statistical point of view, auto-covariance function plays a role of capturing the stochastic pattern of a given signal, which refers to the time dependence structure. Therefore, our approach is not on extracting deterministic pattern in original time domain. This makes our approach to be significantly different from the sparse representation of signals or time frequency decomposition methods. Instead, we aim at first extracting the key signal features, by transforming them to an auto-covariance function, so that the periodogram can be obtained. Notice that the periodogram $I(\omega_j)$ is just the Fourier transform of the auto-covariance function, which captures the quadratic covariation of the lagged signals in the spectral domain. $I(\omega_j)$ is also called a power spectrum of a signal X_t . Because of its definition, the periodogram captures the distribution of covariation of a signal in the spectral domain. The larger is 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, in this work, we will only focus on the analysis of the power spectrum in a sub-interval, i.e. we will analyze only the first 200 frequency values for the given signals.

To illustrate the points stated above, we present a set of results of the power spectra for selected different types of EEG signals in Fig. 1. 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. In particular, we observe that the signal powers for high frequency values are still high for both the signals from the set D and the set E, but they are significantly decayed for both the set B and the set C. This observation inspires us to consider the functional representation of signal power spectra, as the functionality of signals within each set is now significantly enhanced when compared to the original time series plots. However, as we can see from Fig. 1, the power spectra for each set of signals are still very noisy and further smoothing may be required, in order to achieve a better performance on clustering of extracted features.

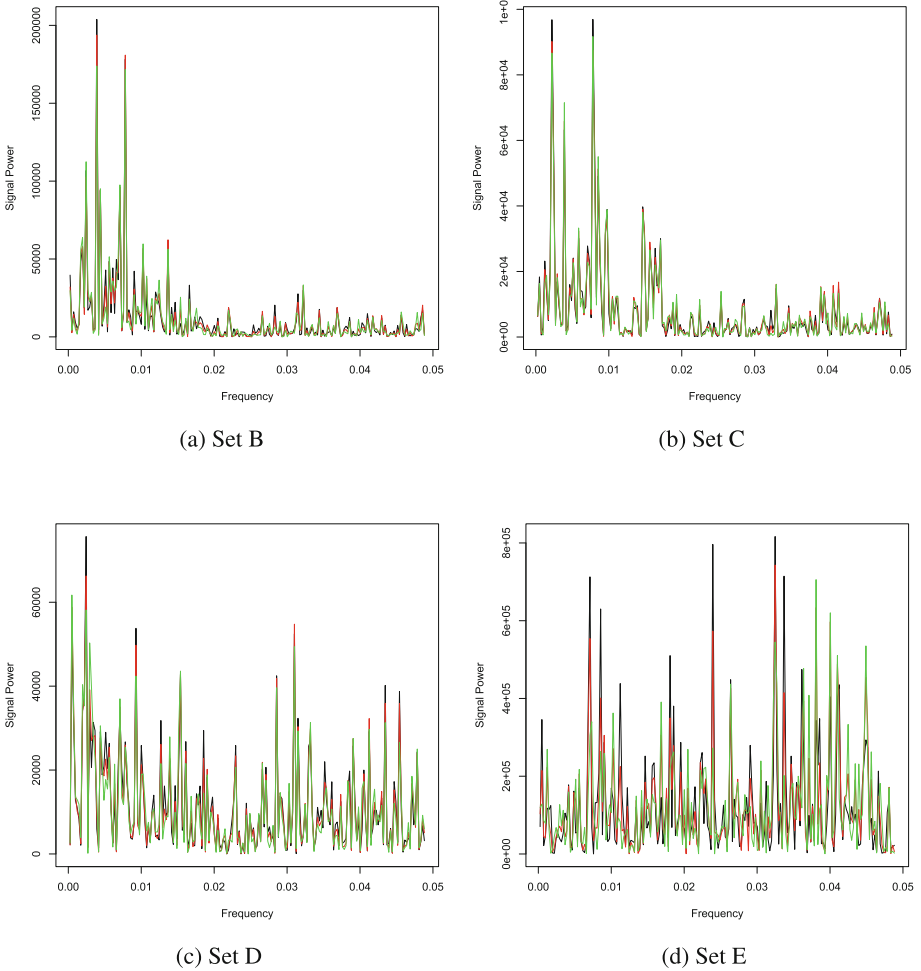


Fig. 1. Sample plots of the power spectra for data sets B (Normal: Eyes Open), C (Non-epileptogenic zone), D (Epileptogenic zone) and E (Seizure onset), respectively. Each plot contains three sample power spectra of the first 200 frequency values. This Figure appears in the proceeding of ICPRAM 2019 [25].

2.2 Functional Representation of Signal Power Spectrum

Our discussion on the signal power spectra in the previous section shows that, they are in fact noisy and further processing is required before we can extract the signal features for clustering. Otherwise, the extracted signal features may be distorted by such noises and a performance of further clustering may not be ideal. Also, a signal is often sampled at a discrete time and further processing is required to make it functional. Equipped with functional data analysis, we can model the power spectrum of a signal with discrete observations by a linear

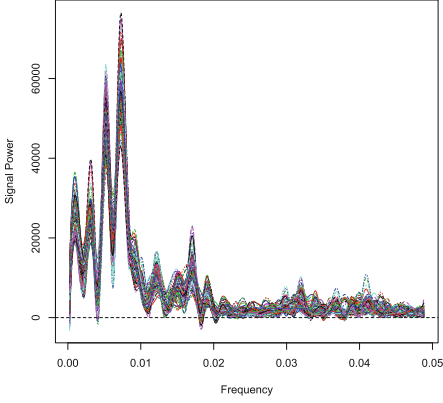
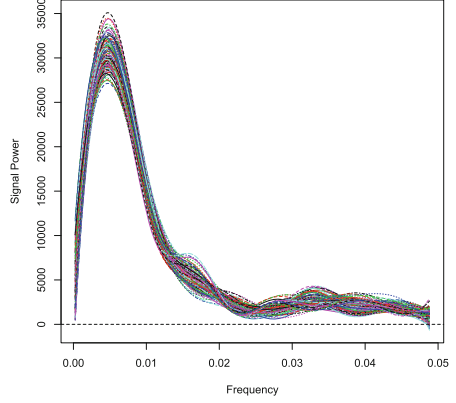
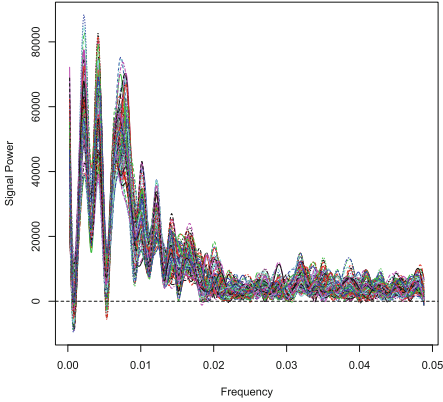
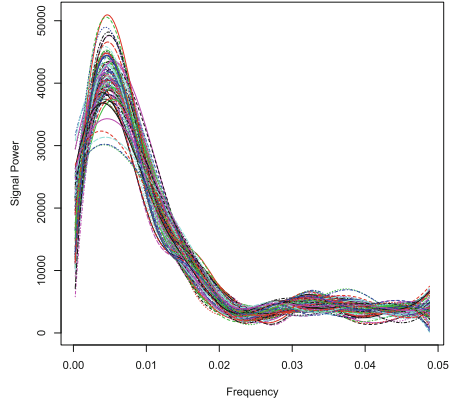
(a) Set A, $K = 50$ (b) Set A, $K = 10$ (c) Set B, $K = 50$ (d) Set B, $K = 10$

Fig. 2. The plots of functional power spectra for data sets A (Normal: Eyes Closed) and B (Normal: Eyes Open), respectively, with different choices of K value. 10 B-splines basis functions are used to smooth sample power spectra.

combination of a set of continuous basis functions. This makes feature extraction of signals to be more mathematically tractable. For a given i th signal, we can expand the power spectrum $I_i(\omega)$ by

$$I_i(\omega) = \sum_{k=1}^K \alpha_{ik} \phi_k(\omega), \quad (3)$$

where ω is the frequency value, α_{ik} is the coefficient of the k th basis function and K is the total number of basis functions. Since our objective is not to fully represent a power spectrum using a set of functional basis, K is assumed to be

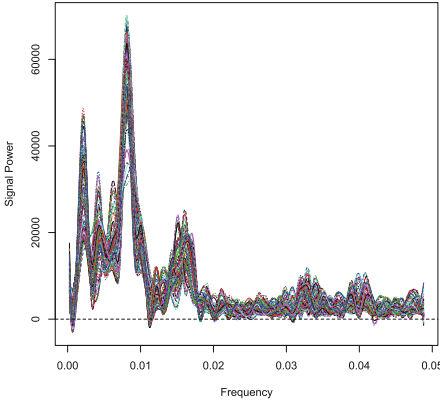
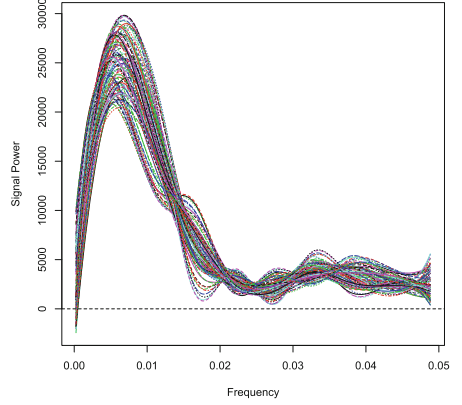
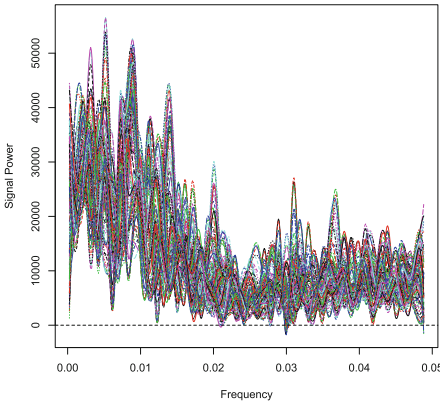
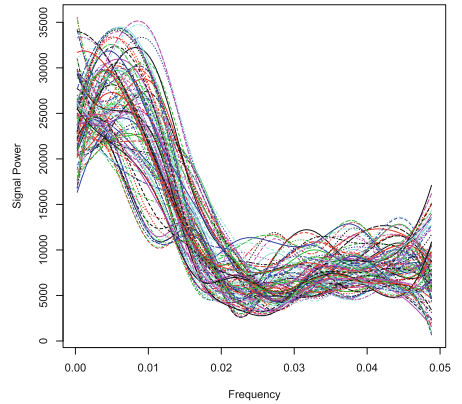
(a) Set C, $K = 50$ (b) Set C, $K = 10$ (c) Set D, $K = 50$ (d) Set D, $K = 10$

Fig. 3. The plots of functional power spectra for data sets C (Non-epileptogenic zone) and D (Epileptogenic zone), respectively, with different choices of K value. 10 B-splines basis functions are used to smooth sample power spectra.

finite and relatively small. That is, we approximate the power spectrum by a linear combination of K basis functions, but for the ease of explanation, we use an equal sign in Eq. (3). However, from the computational perspective, K being finite is always true, as the total number of basis functions needed to approximate the power spectrum is less than the total number of frequency values that we focus on. Notice that, within the discussion of this section, we do not isolate the mean function from the representation of signal. In the later discussion, we will separate the mean function from the signal expansion because of the need of studying functional variation.

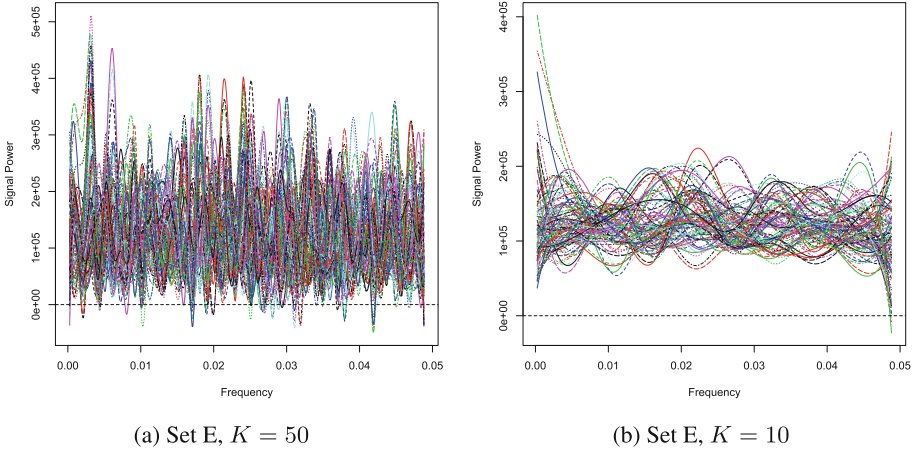


Fig. 4. The plots of functional power spectra for data sets E (Epileptic Seizure), with different choices of K value. 10 B-splines basis functions are used to smooth sample power spectra.

When considering a sample of N signals, that is $i = 1, 2, \dots, N$, in the above Eq. (3), the matrix notation of these power spectra becomes

$$\mathbf{I}(\omega) = \mathbf{A}\phi(\omega), \quad (4)$$

where $\mathbf{I}(\omega) = [I_1(\omega), I_2(\omega), \dots, I_N(\omega)]^\top$ is a column vector of length N and $\phi(\omega) = [\phi_1(\omega), \phi_2(\omega), \dots, \phi_K(\omega)]^\top$ is a column vector of length K containing the basis functions. \mathbf{A} is the coefficient matrix of the size $N \times K$, i.e.

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1K} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \dots & \alpha_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{N1} & \alpha_{N2} & \alpha_{N3} & \dots & \alpha_{NK} \end{bmatrix}. \quad (5)$$

Notice that, the set of basis functions $\phi(\omega)$ can be different for various groups of signals, however, given the fact that we will be considering signals that share many commonalities in the spectral domain, it is reasonable to use the same set of basis functions. This allows us to extract signal features within the same feature space. So, in this work, we will hypothesize that signals can be represented by using the same set of basis functions.

2.3 Functional Probes

The descriptive methods such as functional mean, functional variance or functional covariance allow us to see functional central tendency and functional variation patterns. However, they may be functional and high-dimensional as well.

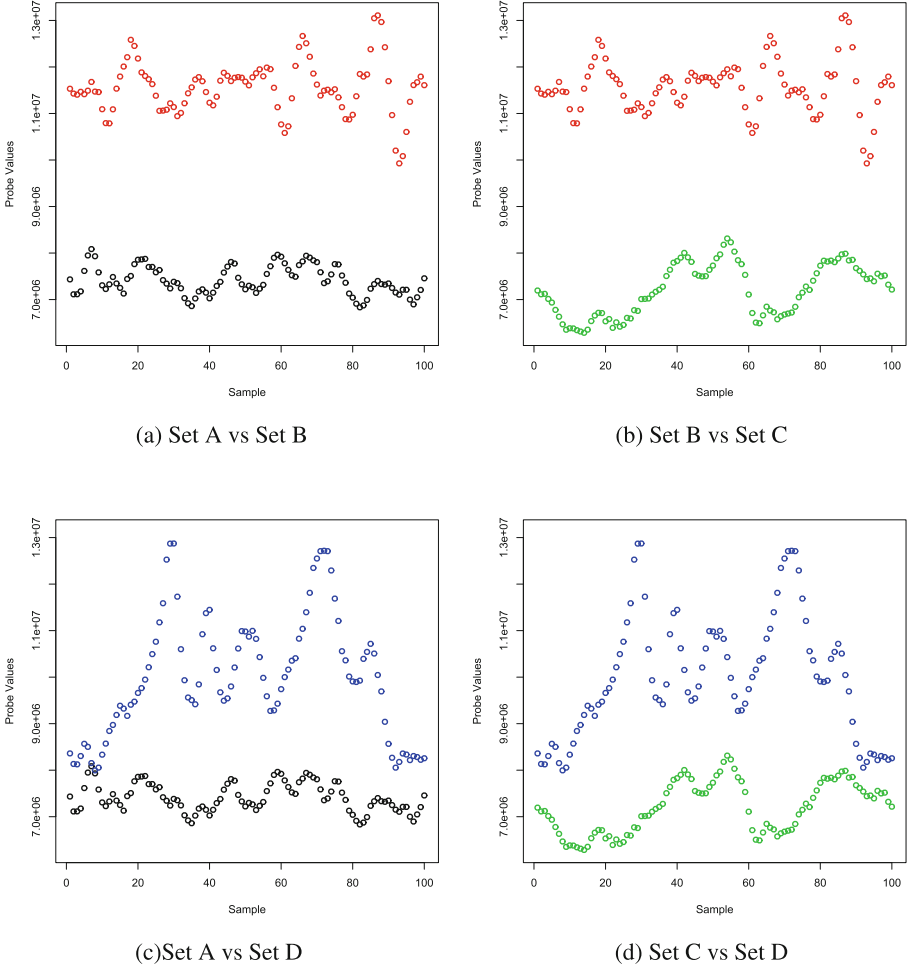


Fig. 5. The plots of extracted probe values using mean function of power spectrum of Set D as the functional probe for data sets A (Normal: Eyes Closed, Black), B (Normal: Eyes Open, Red), C (Non-epileptogenic zone, Green) and D (Epileptogenic zone, Blue), respectively. (Color figure online)

If this is the case, for classification purpose, a dimension of these functional descriptive statistics needs to be further reduced. As a possible dimension reduction method an application of functional probes may be considered. A probe ρ_ξ is a measure that allows us to see specific variation by defining a functional weight $\xi(\omega)$ and it is defined as the inner product of functions $\xi(\omega)$ and $I(\omega)$.

$$\rho_\xi = \int \xi(\omega)I(\omega)d\omega. \quad (6)$$

The $\xi(\omega)$ has to be structured so that we can extract specific features or meaningful patterns of the variation in power spectrum $I(\omega)$. The probe values for i th signal power spectrum using functional mean and functional standard deviation of the j th group power spectrum are defined, respectively, as

$$\rho_{\bar{I}_{ij}} = \int \bar{I}^{(j)}(\omega) I_i(\omega) d\omega = \sum_{k_1=1}^K \sum_{k_2=1}^K \alpha_{ik_1} \bar{\alpha}_{k_2}^{(j)} \int \phi_{k_1}(\omega) \phi_{k_2}(\omega) d\omega, \quad (7)$$

$$\rho_{S_{ij}} = \int S_{I_i(\omega)}^{(j)} I_i(\omega) d\omega. \quad (8)$$

The functional probe values capture the similarity between the weight function and the i th power spectrum of a signal. Using functional mean as a probe, when the basis functions are orthogonal, i.e., $\int \phi_{k_1}(\omega) \phi_{k_2}(\omega) d\omega = 0$, for $k_1 \neq k_2$, and $\int \phi_k^2(\omega) d\omega = 1$, for $k = 1, 2, \dots$, the probe values become

$$\rho_{\bar{I}_{ij}} = \sum_{k=1}^K \alpha_{ik} \bar{\alpha}_k^{(j)}, \quad (9)$$

and they can be interpreted as a similarity measure between two different groups of signals in the spectrum domain, in terms of overall pattern. Unlike the case of using functional mean as a probe, the closed form does not exist for the probe values using functional standard deviation. In this work, we continue to investigate how the extracted signal features (i.e., probe values) behave when functional mean and functional standard deviation are used as the functional probes. However, we focus only on one dimensional quantity. For two dimensional patterns, we refer to [25].

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. by the variance-covariance function, then the functional probe value becomes

$$\int \xi(\omega_2) v(\omega_1, \omega_2) d\omega_2. \quad (10)$$

This is exactly the left hand side of eigenequation for solving eigenvalues and eigenvectors in functional principal component analysis. More discussion of this type of probe will be provided later.

2.4 Classical Principal Component Analysis

In multivariate statistics, principal component analysis (PCA) of a p -variate random vector $X = (X_1, X_2, \dots, X_p)$ looks for a set of weight values, denoted by $\xi_j = (\xi_{1j}, \xi_{1j}, \dots, \xi_{pj})$, so that, at the j th step, the variance of linear combinations of variable X_i is maximized. That is,

$$Var\left(\sum_{i=1}^p \xi_{ij} X_i^{(j)}\right) \quad (11)$$

is maximized or in matrix notation, $Var(X^{(j)}\xi_j^\top)$ is maximized, where

$$X^{(j)} = X^{(j-1)} - \xi_{j-1}^\top X^{(j-1)}, \text{ for } j = 1, 2, \dots, p, \quad (12)$$

where $X^{(0)}$ is defined as $\mathbf{0}$. This process is repeated for $j = 1, 2, \dots, p$, and at each step, it is subject to

$$\sum_{i=1}^p \xi_{ij}^2 = 1 \text{ and } \sum_{i=1}^p \xi_{ij}\xi_{il} = 0, \text{ for } j < l, \text{ and } 1 \leq l, j \leq p.$$

An alternative approach of finding a solution of ξ_j is by using a singular value decomposition (SVD) of the data matrix \mathbf{X} , which contains N realizations of X . The solution of ξ_j is the j th eigenvector obtained from the SVD of the data matrix \mathbf{X} . In this work, the data matrix \mathbf{X} becomes $\mathbf{I}(\omega)$, which is a $N \times p$ data matrix, where p is the total number of frequency values being considered. Technically, the SVD of $\mathbf{I}(\omega)$ is a factorization of the form $\mathbf{U}\Sigma\mathbf{V}$, where \mathbf{U} is a $N \times N$ unitary matrix, Σ is a $N \times p$ diagonal 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 eigenvectors and right eigenvectors 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 a problem of computing $\mathbf{I}(\omega)\xi_j^\top$, for $j = 1, 2, \dots, p$. For example, the first principal component scores set is $\mathbf{I}(\omega)\xi_1^\top$, and the second principal component scores set is $\mathbf{I}(\omega)\xi_2^\top$, etc.

We should also realize that the objective function in PCA can be rewritten as $\xi_j \mathbf{X}^\top \mathbf{X} \xi_j^\top$, assuming that vector \mathbf{X} is centred. In this case, $\mathbf{X}^\top \mathbf{X}$ becomes the variance - covariance matrix. The vector ξ_j is still the eigenvector that is associated with the variance and covariance matrix. All of these can help us understand the relationships among the functional probe, PCA and the functional PCA, which will be discussed later. 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 (6) is maximized, subject to the constraint that $\int \xi_l(\omega)\xi_j(\omega)d\omega = 0$ for $l \neq j$, and $\int \xi_j^2(\omega)d\omega = 1$, then it becomes the functional PCA. In this case, the functional probe values are the principal component scores and the weight function becomes functional principal component loadings.

2.5 Functional Principal Component Analysis

So far, we have explained how functional probe and classical principal component analysis work in analyzing the power spectra of random signals. We have also mentioned that when principal component analysis is combined with functional probe, it leads to functional principal component analysis. That is, functional principal component analysis is a typical statistical analysis procedure that aims at maximizing the functional probe values with the use of variance-covariance matrix as an input to functional probe.

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

$$I_i(\omega) = \mu(\omega) + \sum_{k=1}^K \beta_{ik} \phi_k(\omega), \quad (13)$$

where $\mu(\omega)$ is the functional mean of power spectrum. Here, we consider an approximation of $I_i(\omega)$ rather than exact representation of the function using infinite number of basis functions. Of course, we can not rule out the possibility of having a very large K in representing a signal, but in this work, we will consider a sparse representation of the power spectrum. Also, obtaining the functional mean of power spectrum requires an estimate. Without loss of generality, we can simply take the grand mean over all ω , denoted by μ_0 , as an estimate, so that the Eq. (13) becomes

$$I_i(\omega) = \mu_0 + \sum_{k=1}^K \beta_{ik} \phi_k(\omega), \quad (14)$$

or in a matrix notation

$$\mathbf{I} - \mu = \mathbf{C}\phi, \quad (15)$$

where $\mu = (\mu_0, \mu_0, \dots, \mu_0)^\top$, $\phi = (\phi_1, \phi_2, \dots, \phi_K)^\top$, and the coefficient matrix \mathbf{C} is $N \times K$ that can be written as a matrix:

$$\mathbf{C} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \dots & \beta_{1K} \\ \beta_{21} & \beta_{22} & \beta_{23} & \dots & \beta_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{N1} & \beta_{N2} & \beta_{N3} & \dots & \beta_{NK} \end{bmatrix}.$$

Now, we will describe how to obtain the function principal components and their scores. First, let us denote the variance-covariance function of the power spectra \mathbf{I} by $v(\omega_1, \omega_2)$. In a matrix notation this function is defined as

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

To solve for the eigenfunction, first we have to solve the following eigenequation for an appropriate eigenvalue λ

$$\int v(\omega_1, \omega_2) \xi(\omega_2) d\omega_2 = \lambda \xi(\omega_1). \quad (16)$$

Suppose that the eigenfunction $\xi(\omega)$ has an expansion

$$\xi(\omega) = \sum_{k=1}^K b_k \phi_k(\omega), \quad (17)$$

or in matrix notation

$$\xi(\omega) = \phi^\top(\omega) \mathbf{b}, \quad (18)$$

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

$$\begin{aligned} \int v(\omega_1, \omega_2) \xi(\omega_2) d\omega_2 &= \frac{1}{N-1} \int \phi^\top(\omega_1) \mathbf{C}^\top \mathbf{C} \\ &\quad \phi(\omega_2) \phi^\top(\omega_2) \mathbf{b} d\omega_2. \\ &= \frac{1}{N-1} \phi^\top(\omega_1) \mathbf{C}^\top \mathbf{C} \Phi \mathbf{b}, \end{aligned} \quad (19)$$

where $\Phi = \int \phi(\omega) \phi^\top(\omega) d\omega$ is the matrix that is given as follows:

$$\Phi = \begin{bmatrix} \int \phi_1 \phi_1 d\omega & \int \phi_1 \phi_2 d\omega & \dots & \int \phi_1 \phi_K d\omega \\ \int \phi_2 \phi_1 d\omega & \int \phi_2 \phi_2 d\omega & \dots & \int \phi_2 \phi_K d\omega \\ \vdots & \vdots & \ddots & \vdots \\ \int \phi_K \phi_1 d\omega & \int \phi_K \phi_2 d\omega & \dots & \int \phi_K \phi_K d\omega \end{bmatrix}.$$

The eigenequation in (16) becomes

$$\frac{1}{N-1} \phi^\top(\omega) \mathbf{C}^\top \mathbf{C} \Phi \mathbf{b} = \lambda \phi^\top(\omega) \mathbf{b}. \quad (20)$$

Since (20) must hold for all ω , this implies that the following matrix equation must hold

$$\frac{1}{N-1} \mathbf{C}^\top \mathbf{C} \Phi \mathbf{b} = \lambda \mathbf{b}. \quad (21)$$

To obtain the required principal components, we define $\mathbf{u} = \Phi^{1/2} \mathbf{b}$ and the Eq. (21) becomes

$$\frac{1}{N-1} \Phi^{1/2} \mathbf{C}^\top \mathbf{C} \Phi^{1/2} \mathbf{u} = \lambda \mathbf{u}. \quad (22)$$

By solving the symmetric eigenvalue problem for \mathbf{u} in (22), and then computing $\mathbf{b} = \Phi^{-1/2} \mathbf{u}$ to get the eigenfunction $\xi(\omega)$, we get that

$$\xi(\omega) = \phi^\top(\omega) \Phi^{-1/2} \mathbf{u}. \quad (23)$$

Solving the eigenvalue problem in (22) will produce K different eigenfunctions and their corresponding eigenvalues which we denote by $(\lambda_j, \xi_j(\omega))$.

If $\phi_k(\omega)$ are orthonormal, then Φ becomes the $K \times K$ identity matrix and the eigenanalysis of the functional PCA problem in (21) reduces to

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

which is the multivariate PCA that replaces variance-covariance matrix by the coefficient matrix \mathbf{C} obtained from the function approximation of power spectrum. However, one should realize that this is not a standard multivariate PCA. From the discussion above, a multivariate PCA conducts eigenanalysis for a $p \times p$ covariance matrix. With function approximation using K basis functions, the eigenanalysis of functional PCA is applied to $K \times K$ coefficient matrix, which depends on the value of K . In case of sparse approximation, which gives a small value of K , solving an eigenequation is much more efficient from the point of view of computational complexity.

2.6 B-Spline as Functional Basis

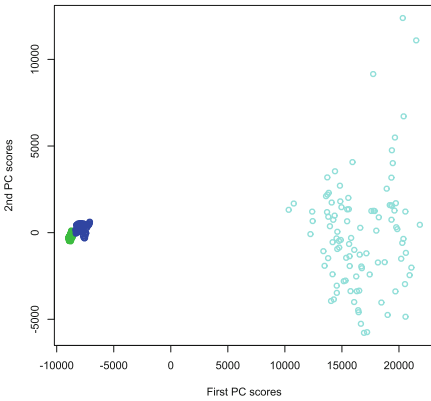
In this work, we choose the spline function as functional basis which is called B-spline [3, 19]. B-spline has been used very often in numerical analysis to approximate a function or a surface. Assume that there is a sequence of non-decreasing real numbers (i.e., $t_k \leq t_{k+1}$), such that $t_0 \leq t_1 \leq \dots \leq t_{N-1}$, and N is the length of a signal. We call the set $\{t_k \mid k \in \mathbb{Z}\}$ a knot set and each value t_k is referred to as a knot. Next, we define the augmented knot set $t_{-v+2} = \dots = t_0 \leq t_1 \leq \dots \leq t_{N-1} = t_N \dots = t_{N+v-2}$, where v is the order of the B-spline. We have appended the lower and upper indexes because of the use of recursive formula of the B-spline. Furthermore, we reset the index, so that the new index in the augmented knot set becomes $k = 0, 1, \dots, N + 2v - 3$. For each augmented knot t_k , $k = 0, 1, \dots, N + 2v - 3$, we define B-spline recursively as follows:

$$B_{k,0}(x) = \begin{cases} 1, & \text{if } t_k \leq x < t_{k+1} \\ 0, & \text{otherwise} \end{cases}$$

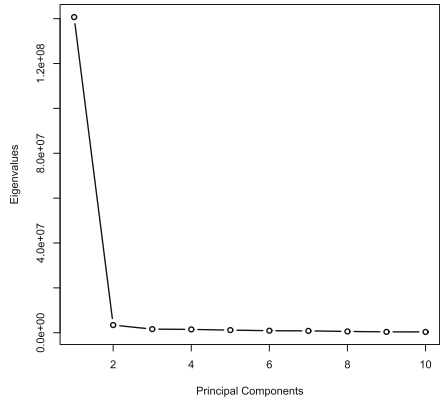
$$B_{k,j}(x) = \gamma_{k,j+1}(x)B_{k,j}(x) + [1 - \gamma_{k+1,j+1}]B_{k+1,j}(x), \quad (24)$$

for $j = 0, 1, \dots, v - 1$, where

$$\gamma_{k,j+1}(x) = \begin{cases} \frac{x-t_k}{t_{k+j}-t_k}, & \text{if } t_{k+j} \neq t_k \\ 0, & \text{otherwise} \end{cases}.$$



(a) PC scores



(b) Eigenvalues

Fig. 6. The results of principal component scores and eigenvalues of power spectra for C (Non-epileptogenic zone, Green), D (Epileptogenic zone, Blue) and E (Epileptic seizure onset, Light Blue). (Color figure online)

After this recursive process, the maximum number of non-zero basis functions is N , and they are denoted by $B_{1,v-1}, B_{1,v-1}, \dots, B_{N,v-1}$, and they are called

B-spline basis functions of $v - 1$ degree. If the number of interior points of the knot set, denoted by N^* , is smaller than $N - 2$, then the total number of basis functions is $K = N^* + v$. When $v = 4$, these functions are called cubic B-spline basis functions. In this work, we assume

$$\phi_k(a) = B_{k,3}(a), \text{ for } k = 1, 2, \dots, N^* + 4.$$

So, the maximum number of basis functions used in this work is $K = N^* + 4$.

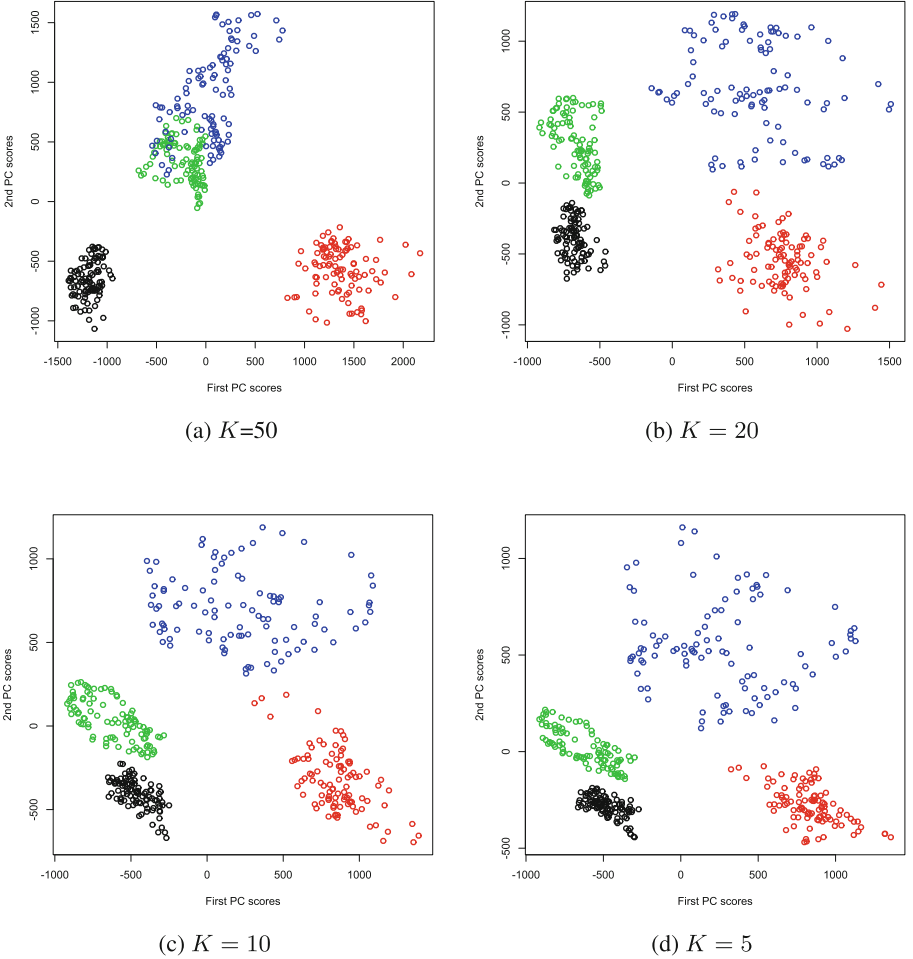


Fig. 7. The evolution of extracted first two principal component scores of power spectra under different choices of K for data sets A (Normal: Eyes Closed, Black), B (Normal: Eyes Open, Red), C (Non-epileptogenic zone, Green) and D (Epileptogenic zone, Blue). (Color figure online)

2.7 Feature Extraction by Functional Principal Component Analysis

After the j th eigenfunction $\xi_j(\omega)$ is obtained, we can extract the principal component scores, denoted by P_j , for the given power spectrum $\mathbf{I}(\omega)$ as follows

$$P_j = \int \mathbf{I}(\omega)\xi_j(\omega)d\omega, \quad j = 1, \dots, K. \quad (25)$$

Substituting (15) and (23) to the equation above, we get

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

Thus, P_1 is the first principal component score vector of the N power spectra of signals, and P_2 is the second principal component score vector, and so on. One of the strengths of using PCA is retaining a small number of principal components, so that, the dimension of feature vector is low. Using the functional PCA on the Fourier power spectra, we have been able to focus on only the first two PCs for clustering the EEG signals.

3 Results

In this study, we use the same data set as in [25], which is from the University of Bonn, Germany (http://epileptologie-bonn.de/cms/front_content.php?idcat=193), but we extend our study by including the set E. We focus on both the epilepsy diagnosis and the epilepsy seizure detection problems. From the results displayed in Figs. 2 and 3, one can see that the numbers of basis functions have a significant effect on the shape of the power spectra. When $K = 10$ is used to smooth out the power spectra, then the smoothed signal power spectra behave similarly for both types of signals, one coming from healthy people (sets A and B) and another one coming from patients for which the signals 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. These differences become more clear when the number of basis functions is $K = 50$. It is understood that when the values of K increase, the smoothed power spectra tend to capture more of the local patterns. This makes a significant difference between different sets of signals. Furthermore, the patterns of power spectra associated with signals from patients with seizure onset (set E)

are completely different from other types of signals. From Fig. 4, one can see that there is no clear functionality in the power spectra and with the decrease of K values the overall pattern among the power spectra is still not recognizable. This may suggest the ease of clustering this type of signals from the others. Overall, the graphical display offers some evidence that suitable clustering methods may differentiate these types of signals successfully. In particular, we can aim for the clustering between the sets A and B, this allows us to see the differences caused by artifacts. We can also combine the signals from healthy people, i.e. the sets A and B, and combine the signals from patient without seizure onset, i.e. the sets C and D, in order to see if there is a clustering effect between patients' EEGs and healthy people's EEGs. However, one question still remains. Should the clustering be based on the one that leads to significant overall difference, or the one that offers big difference in local patterns within the power spectra? This question will be answered later in the analysis of the evolution of extracted features.

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 studied. In our study, the best results, in terms of separability of features, are the ones that use the functional mean calculated from the set D. Using the functional mean as a probe, we extract a single-dimensional feature vector from a given signal power spectrum. Figure 5 clearly display the pattern, which shows a great separability of extracted features (i.e., functional probe values), due to the dimension reduction. We observe that the artifacts (i.e., eyes open or eyes closed) associated with the healthy people can be identified when mapping the power spectra of signals on the power spectra of epileptic signals (i.e., set D signals). Using only a single dimension of features, the separation of signals can be achieved for the set A and the set B. Similarly, these single dimension features are highly separable for other cases. This may suggest that the functional probe approach has certain merits in automatic clustering of different types of EEG signals.

To demonstrate the application to epileptic seizure detection, we use functional PCA to extract the principal components of the power spectra of signals from the sets C, D and E. From the results one can see if there exist some dominant signal components. Additionally, the corresponding scores can be used for clustering. The obtained results are displayed in Fig. 6. From Fig. 6(a), one can see that the extracted features for non-seizure signals (sets C and D) and seizure signals (set E) form into clusters and they are linearly separable in first PC. Figure 6(b) display that, the first eigenvalue is very dominant, which explains why the extracted features are highly separable in the first PC. This implies that the extracted first PC scores can be successfully classified using a simple classification method, such as k-NN. The second PC scores are not helpful in contributing to the cluster effect as they are completely overlapped. We further investigate the effect of the number of basis functions (i.e., K) on the sepa-

rability of extracted signal features (i.e., principal component scores of power spectra). The proposed method is applied to signals from sets A, B, C and D only by varying the value of K . The obtained results are displayed in Fig. 7. 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 $K = 50$, 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 20. When $K = 5$, the obtained results are considered to be optimal in the sense of feature separability. However, the overall separability between healthy and epileptic signals is not affected by the number of basis functions. This may suggest that applying the proposed methods to the epilepsy diagnosis problem can be a successful tool, which allows separating the healthy and patient signal.

4 Concluding Remarks

Clustering and classification of high-dimensional data are important aspects in both pattern recognition and artificial intelligence. To be successful in dealing with clustering and classification, dimension reduction of high-dimensional data plays an important role. In this work, we focus on the study of using feature extraction as a dimension reduction approach. We use EEG signals to illustrate our proposed methodologies. We first transform EEG signals to the spectral domain to obtain their power spectra. Next, we apply functional principal component analysis, which is considered to be a special case of functional probe, to further investigate the characteristics of the signals and to extract their features for clustering. We have demonstrated that functional principal component analysis in spectral domain is useful for better understanding of different types of EEG signals. Furthermore, the extracted features can be used for signal classification. We have also investigated the effect of sparsity on the performance of separating signal features. We have observed that the separability of extracted features is significantly improved, when the number of signal components used for approximating the power spectra decreases. This may imply that the sparse approximation for signal approximation in spectral domain is a necessary step for better performance of signal clustering. From an application perspective, the obtained results demonstrate that the proposed method may be useful for both epilepsy diagnosis and epileptic seizure detection. Future work will focus on the study of wavelet spectral domain functional PCA and its application to clustering general random signals such as financial time series.

References

1. Alickovic, E., Kevric, J., Subasi, A.: Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomed. Signal Process. Control* **39**, 94–102 (2018)
2. Bouveyron, C., Girard, S., Schmid, C.: High-dimensional data clustering. *Comput. Stat. Data Anal.* **52**(1), 502–519 (2007)
3. De Boor, C.: *A Practical Guide to Splines*, vol. 27. Springer, New York (1978)
4. Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K., Hamdan, H.: A machine learning system for automated whole-brain seizure detection. *Appl. Comput. Inf.* **12**(1), 70–89 (2016)
5. Gandhi, T., Panigrahi, B.K., Anand, S.: A comparative study of wavelet families for EEG signal classification. *Neurocomputing* **74**(17), 3051–3057 (2011)
6. Garcia, G.N., Ebrahimi, T., Vesin, J.M.: Support vector EEG classification in the Fourier and time-frequency correlation domains. In: *First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings*, pp. 591–594. IEEE (2003)
7. Jimenez, L.O., Landgrebe, D.A.: Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **28**(1), 39–54 (1998)
8. Kriegel, H.P., Kröger, P., Zimek, A.: Clustering high-dimensional data: a survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Trans. Knowl. Discov. Data (TKDD)* **3**(1), 1 (2009)
9. Li, D., Pedrycz, W., Pizzi, N.J.: Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification. *IEEE Trans. Biomed. Eng.* **52**(6), 1132–1139 (2005)
10. Liang, S.F., Wang, H.C., Chang, W.L.: Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. *EURASIP J. Adv. Signal Process.* **2010**, 62 (2010)
11. Lima, C.A., Coelho, A.L., Chagas, S.: Automatic EEG signal classification for epilepsy diagnosis with relevance vector machines. *Expert Syst. Appl.* **36**(6), 10054–10059 (2009)
12. Nyan, M., Tay, F., Seah, K., Sitoh, Y.: Classification of gait patterns in the time-frequency domain. *J. Biomech.* **39**(14), 2647–2656 (2006)
13. Phinyomark, A., Phukpattaranont, P., Limsakul, C.: Feature reduction and selection for EMG signal classification. *Expert Syst. Appl.* **39**(8), 7420–7431 (2012)
14. Qazi, K.I., Lam, H., Xiao, B., Ouyang, G., Yin, X.: Classification of epilepsy using computational intelligence techniques. *CAAI Trans. Intell. Technol.* **1**(2), 137–149 (2016)
15. Ramsay, J.: *Functional Data Analysis*. Encyclopedia of Statistics in Behavioral Science (2005)
16. Ramsay, J.O., Silverman, B.W.: *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York (2007)
17. Subasi, A., Gursoy, M.I.: EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Syst. Appl.* **37**(12), 8659–8666 (2010)
18. Truong, N.D., Kuhlmann, L., Bonyadi, M.R., Yang, J., Faulks, A., Kavehei, O.: Supervised learning in automatic channel selection for epileptic seizure detection. *Expert Syst. Appl.* **86**, 199–207 (2017)

19. Unser, M., Aldroubi, A., Eden, M., et al.: B-spline signal processing: part I theory. *IEEE Trans. Signal Process.* **41**(2), 821–833 (1993)
20. Wang, J.L., Chiou, J.M., Müller, H.G.: Functional data analysis. *Ann. Rev. Stat. Appl.* **3**, 257–295 (2016)
21. Xie, S., Krishnan, S.: Signal decomposition by multi-scale PCA and its applications to long-term EEG signal classification. In: *The 2011 IEEE/ICME International Conference on Complex Medical Engineering*, pp. 532–537. IEEE (2011)
22. Xie, S., Krishnan, S.: Wavelet-based sparse functional linear model with applications to EEGs seizure detection and epilepsy diagnosis. *Med. Biol. Eng. Comput.* **51**(1–2), 49–60 (2013)
23. Xie, S., Krishnan, S.: Dynamic principal component analysis with nonoverlapping moving window and its applications to epileptic EEG classification. *Sci. World J.* **2014**, 1–10 (2014)
24. Xie, S., Krishnan, S.: Model based sparse feature extraction for biomedical signal classification. *Int. J. Stat. Med. Res.* **6**(1), 10–21 (2017)
25. Xie, S., Lawniczak, A.: Feature extraction of eeg in spectral domain via functional data analysis. In: *International Conference on Pattern Recognition Applications and Methods*, pp. 118–127 (2019)
26. Xie, S., Lawniczak, A.T., Song, Y., Liò, P.: Feature extraction via dynamic PCA for epilepsy diagnosis and epileptic seizure detection. In: *2010 IEEE International Workshop on Machine Learning for Signal Processing*, pp. 337–342. IEEE (2010)
27. Yu, L., Liu, H.: Feature selection for high-dimensional data: a fast correlation-based filter solution. In: *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 856–863 (2003)
28. Zhang, Z., Xu, Y., Yang, J., Li, X., Zhang, D.: A survey of sparse representation: algorithms and applications. *IEEE Access* **3**, 490–530 (2015)