

Multivariate Time-Series Analysis Via Manifold Learning

Pedro Luiz Coelho Rodrigues, Marco Congedo, Christian Jutten

▶ **To cite this version:**

Pedro Luiz Coelho Rodrigues, Marco Congedo, Christian Jutten. Multivariate Time-Series Analysis Via Manifold Learning. SSP 2018 - 2018 IEEE Workshop on Statistical Signal Processing, Jun 2018, Fribourg-en-Brisgau, Germany. 10.1109/SSP.2018.8450771 . hal-01868167

HAL Id: hal-01868167
https://hal.science/hal-01868167

Submitted on 5 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MULTIVARIATE TIME-SERIES ANALYSIS VIA MANIFOLD LEARNING

Pedro Luiz Coelho Rodrigues, Marco Congedo, and Christian Jutten

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

ABSTRACT

This paper presents a data-driven approach for analyzing multivariate time series. It relies on the hypothesis that high-dimensional data often lie on a low-dimensional manifold whose geometry may be revealed using manifold learning techniques. We define a notion of distance between multivariate time series and use it to determine a low-dimensional embedding capable of describing the statistics of the signals at hand using just a few parameters. We illustrate our method on two simulated examples and two real datasets containing electroencephalographic recordings (EEG).

***Index Terms*—** Manifold Learning, Information Geometry, EEG Signal Analysis

1. INTRODUCTION

The age of big data has changed the way of doing experimental science, making it easier and cheaper than ever to record physical phenomena simultaneously from several different sensors. However, collecting more data does not necessarily mean having more useful information available, since high-dimensional samples often carry a lot of redundancy. One clear example is the case of electroencephalographic (EEG) recordings, where sensors located at close positions of the scalp of a subject tend to record time series which are very correlated to one another.

Because of this redundancy in information, it is common to assume that samples of a dataset with n dimensions (n large) actually live in a low-dimensional manifold $\mathcal{M}$ embedded in $\mathbb{R}^n$ [1]. Such intrinsic low-dimension may then be explored for improving performance in classification, clearer interpretations from data, etc..

Many algorithms have been proposed to reveal the geometry of $\mathcal{M}$ and let us work with a reduced-dimension version of the samples. For instance, linear methods such as principal component analysis (PCA) and multi-dimensional scaling (MDS) [2] assume that the low-dimensional manifold is a hyperplane, while Laplacian eigenmaps [3] and diffusion maps [1] are non-linear methods (often called manifold learning techniques) that try to approximate the eigenfunctions of the Laplace-Beltrami operator over $\mathcal{M}$. All such methods rely on the definition of a notion of similarity between points of the dataset and an optimization problem where one wants to minimize the changes in similarities between each pair of points in the original space and the embedded one.

Recently, there has been some interest in the study of time series via manifold learning techniques, where the goal is to describe changes in a time series' statistics using just a few dimensions. In [4], the authors analysed such changes using a sliding-window approach, where each short interval of time was described according to some features of interest. They measured the similarity between different windows of data with an adapted version of the Mahalanobis distance and used the diffusion maps algorithm to embed the samples into a low-dimensional space.

In this work, we introduce an unsupervised method for analyzing multivariate time series based on the diffusion maps algorithm, with a special emphasis on EEG recordings. We consider two settings, one where the samples of the time series are all available in short trials and each corresponds to one particular state (e.g., 30s of EEG during one sleep stage) and one where we track the changes in the signals' statistics along time (e.g., switch between open and closed eyes every one second).

Our main contribution is to endow the diffusion maps algorithm with a notion of distance between multivariate time series based on well-known results of statistical signal processing [5] and information geometry [6]. For this, we use Riemannian geometry to determine the distances between features representing the statistics of the EEG signals at hand.

2. MATERIALS AND METHODS

2.1. Distances between multivariate time series

Let $\boldsymbol{x}(n)$ denote a Gaussian wide-sense stationary (G-WSS) N-dimensional time series. All the statistical information of $\boldsymbol{x}(n)$ is described by its mean vector and $N \times N$ auto-covariance matrices, defined as:

$$R_{\boldsymbol{xx}}(\tau) = \mathbb{E}\big[\boldsymbol{x}(n)\boldsymbol{x}(n-\tau)^T\big] \quad (\tau \in \mathbb{Z}), \tag{1}$$

or, equivalently, by its cross-spectral density matrices:

$$S_{\boldsymbol{xx}}(f) = \sum_{k=-\infty}^{+\infty} R_{\boldsymbol{xx}}(k) e^{-j2\pi fk}, \tag{2}$$

This research was partly supported by the ERC Advanced Grant CHESS (2012-ERC-AdG-320684).

where $f \in [0, 1]$ is the normalized frequency [5].

When comparing two zero-mean G-WSS time series $\boldsymbol{x}_1(n)$ and $\boldsymbol{x}_2(n)$, one can define a distance based on the description of their statistics via auto-covariance or cross-spectrum matrices. This approach comes from Information Geometry and is more meaningful than simply comparing the samples of a realization of the time series, since it considers the actual random processes that generated them [6].

In this work, we use two distances based on such premise:

- Distance based on the covariance matrices of $\boldsymbol{x}_i(n)$:

$$d_C^2\big(\boldsymbol{x}_1(n), \boldsymbol{x}_2(n)\big) = \delta_R^2(C_1, C_2), \tag{3}$$

  where $C_i = \mathbb{E}\big[\boldsymbol{x}_i(n)\boldsymbol{x}_i(n)^T\big]$ for $i \in \{1, 2\}$ and $\delta_R^2(A, B)$ is the geodesic distance between two symmetric positive definite (SPD) matrices A and B [7]:

$$\delta_R^2(A, B) = \left\| \log\left(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}\right) \right\|_F^2. \tag{4}$$

  Distance (3) is also the distance induced by Fisher's metric on the statistical manifold of normal multivariate random variables with zero mean [6].

- Distance based on the cross-spectrum of $\boldsymbol{x}_i(n)$:

$$d_S^2\big(\boldsymbol{x}_1(n), \boldsymbol{x}_2(n)\big) = \sum_{f_k \in \mathcal{F}} \delta_R^2\big(S_1(f_k), S_2(f_k)\big), \tag{5}$$

  where $S_i(f_k)$ is the cross-spectrum of $\boldsymbol{x}_i(n)$ at $f = f_k$, and $\mathcal{F}$ denotes a set of frequencies of interest.

When working with multivariate EEG recordings, each coordinate of $\boldsymbol{x}(n)$ corresponds to the activity recorded on one electrode positioned on a subject's scalp. We expect the statistics of $\boldsymbol{x}(n)$ to change according to the tasks performed by the subject and assume that they can be tracked via its auto-covariance or cross-spectrum at each time (we always consider $\boldsymbol{x}(n)$ to be zero-mean).

2.2. Diffusion maps

The method of diffusion maps relies on the construction of a pairwise affinity matrix of a set of points $\mathcal{X} = \{X_1, \ldots, X_K\}$, whose (i, j)-th element is given by:

$$\boldsymbol{W}_{ij} = \exp\left(-\frac{d^2(X_i, X_j)}{\epsilon}\right), \tag{6}$$

where $d(\cdot, \cdot)$ is a notion of similarity between two points in $\mathcal{X}$ (usually a measure of distance) and ϵ is the scale of the Gaussian kernel (which can be set as the median value of all the pairwise distances in $\mathcal{X}$ [1]).

Let $\boldsymbol{D}$ be a diagonal matrix whose elements are the sums of rows of $\boldsymbol{W}$, and $\boldsymbol{W}^{\text{norm}} = \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{W} \boldsymbol{D}^{-\frac{1}{2}}$. The spectral decomposition of $\boldsymbol{W}^{\text{norm}}$ gives a sequence of eigenvectors $\varphi_j \in \mathbb{R}^K$ and eigenvalues λ_j such that

$$|\lambda_0| \geq |\lambda_1| \geq \cdots \geq |\lambda_{K-1}|,$$

where $\boldsymbol{W}^{\text{norm}} \varphi_j = \lambda_j \varphi_j$.

One can show that λ_0 is always equal to 1 and φ_0 is the diagonal of $\boldsymbol{D}^{1/2}$ [1]. The next d eigenvectors define a map of the elements in $\mathcal{X}$ to a d-dimensional Euclidean space

$$X_i \mapsto \Phi(X_i) = (\varphi_1(i), \ldots, \varphi_d(i)), \tag{7}$$

where $\varphi_k(i)$ denotes the i-th element of φ_k. This mapping is what we call *spectral embedding*. The dimension d is chosen according to the decay of the eigenvalues of $\boldsymbol{W}^{\text{norm}}$ and is usually much smaller than K. See [1] for more details.

2.3. Spectral embedding of EEG recordings

We consider two applications of spectral embedding via diffusion maps.

In the first case, the dataset is composed of short recordings on N electrodes at different moments of time. We denote the data points by $X_k \in \mathbb{R}^{N \times T}$, with $k \in \{1, \ldots, K\}$, and assume they each represent T samples of a N-dimensional G-WSS time series $\boldsymbol{x}_k(n)$. We use distances (3) and (5) as measures of similarity in the affinity matrix $\boldsymbol{W}$, where the covariance matrices are estimated from the dataset and the cross-spectrum is obtained using Welch's method [5]. The spectral embedding of the X_k's will map similar points to close positions in the Euclidean space, which amounts to clustering time series associated to the same physiological state.

We also consider how the statistics of a multivariate EEG time series $\boldsymbol{x}(n)$ evolve in time. We assume that $\boldsymbol{x}(n)$ is G-WSS in short windows containing L points and use a sliding-window procedure to obtain K points $X_k \in \mathbb{R}^{N \times L}$, each associated to a window of L points around a different instant of time. We estimate the auto-covariances and cross-spectra using the samples of each window and embed them into a low-dimensional Euclidean space with diffusion maps. The trajectories identified in the embedded space describe the changes in statistics of $\boldsymbol{x}(n)$ and can be used to study the dynamics of latent variables.

2.4. Numerical illustrations

2.4.1. Simulated data

We simulated data using a three-dimensional multivariate autoregressive process defined by

$$\boldsymbol{s}(n) = \mathbf{A}_1 \boldsymbol{s}(n-1) + \mathbf{A}_2 \boldsymbol{s}(n-2) + \boldsymbol{u}(n), \tag{8}$$

with

$$\mathbf{A}_1 = \text{diag}\Big(2r_0 \cos(2\pi f_0), 2r_1 \cos(2\pi f_1), 2r_2 \cos(2\pi f_2)\Big),$$

$$\mathbf{A}_2 = \text{diag}(-r_0^2, -r_1^2, -r_2^2),$$

where $\text{diag}(a, b, c)$ denotes a 3×3 diagonal matrix with elements a, b, and c. We fixed $(f_0, f_1, f_2) = (0.1, 0.25, 0.40)$ and chose (r_1, r_2, r_3) according to each example. The $\boldsymbol{u}(n)$ time series is a Gaussian white noise with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{uu}} = \mathbf{I}_3$. The actual $\boldsymbol{x}(n)$ time series used in the examples was obtained after three pre-processing steps on $\boldsymbol{s}(n)$:

(1) Divide each signal $s_i(n)$ in $\boldsymbol{s}(n)$ by its standard deviation so to fix the signal variance to one.

(2) Determine a mixing matrix $\mathbf{M}$ with dimensions 3×3.

(3) Generate the mixed signals: $\boldsymbol{x}(n) = \mathbf{M}\boldsymbol{s}(n)$.

In the example portrayed in Fig. 1, we considered three conditions (states) for the signals generated by (8):

$$\begin{aligned} \mathcal{S}_1 &: (r_1, r_2, r_3) = (0.00, 0.95, 0.95) \\ \mathcal{S}_2 &: (r_1, r_2, r_3) = (0.95, 0.00, 0.95) \\ \mathcal{S}_3 &: (r_1, r_2, r_3) = (0.95, 0.95, 0.00) \end{aligned} \tag{9}$$

We generated $K = 150$ trials (50 for each condition) containing $T = 512$ samples each. Matrix $\mathbf{M}$ was the same for all trials and its elements were randomly obtained from a Gaussian distribution with zero mean and unit variance. The example in Fig. 2 contains 2800 samples of a single-trial time series generated with $(r_1, r_2, r_3) = (0.95, 0.95, 0.95)$ and a time-varying mixing matrix

$$\mathbf{M}(n) = m(n)\mathbf{M}_1(n) + \big(1 - m(n)\big)\mathbf{M}_2(n), \tag{10}$$

where

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{M}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

and $m(n)$ is a parameter that toggles between 0 and 1 every 400 samples.

2.4.2. *Real EEG data*

We analyzed two datasets containing physiological recordings from human subjects.

The example in Figure 3 used data available at the Physionet database [9, 10]. It contains recordings from 9 EEG and 2 EOG electrodes on a subject sleeping for approximately 8 hours. The original sampling frequency was 512 Hz but we downsampled it to 128 Hz after band-filtering the EEG signals between 8 Hz and 40 Hz and the EOG signals between 0.05 Hz and 10 Hz. A specialist in sleep data analysis was responsible for cutting the recordings into several 30-sec clips and then classifying them according to which sleep stage (S1, S2, S3, and REM) they belonged. We ended up with a dataset containing $K = 564$ trials of $L = 3840$ samples each.

The example depicted in Fig. 4 used recordings from 16 EEG electrodes of a healthy subject instructed to change between open or closed eyes every eight seconds. The sampling frequency was 128 Hz and was filtered between 8 Hz and 40 Hz. The dataset had a single-trial with 11520 samples (90 seconds).

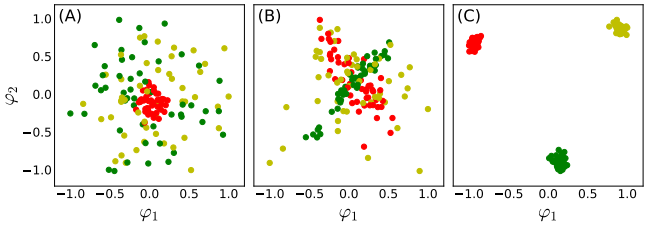


Fig. 1. Spectral embedding of the signals specified by (9) using only two dimensions ($d = 2$). Each point corresponds to a trial $X_i \in \mathbb{R}^{3\times 512}$ and the colors indicate to which condition it is associated ($\mathcal{S}_1$ is yellow, $\mathcal{S}_2$ is red, $\mathcal{S}_3$ is green). We used diffusion maps with three types of distance: (A) Euclidean distance between the X_i matrices, (B) distance defined in (3), and (C) distance defined in (5).

3. RESULTS AND DISCUSSION

We observe in Fig. 1 that choosing an adequate distance between time series is of prime importance, as seen by the lack of a clear pattern in the spectral embedding using the Euclidean distance between time series (Fig.1(A)) or the distance (3) between covariance matrices (Fig.1(B)). This happens because all the X_i's generated in the example have the same covariance matrix for all three states $\mathcal{S}_i$. Only the cross-spectrum is capable of differentiating each state.

It could be argued that the quality of the clustering with each type of distance also depends on the number of eigenvectors considered from the diffusion maps (Fig.1 shows the results with just the first two). To verify this hypothesis, we assessed the quality of the clustering performed by a Gaussian Mixture Model (GMM) [2] on the embedded space considering an increasing number of dimensions. Assuming as true labels the classes given by a specialist, we did not observe any improvement on the modified Rand scores [11] when considering more eigenvectors for any of the distances.

Our next simulated example demonstrates that diffusion maps can be useful for studying the temporal dynamics of the statistics of a time series. Fig. 2 shows the values of the first coordinate of the embedding of a sliding-window and compares them with a latent parameter responsible for changing the statistics of $\boldsymbol{x}(n)$ along time. We estimated only the covariance matrices of the windows of data and used them to embed the time series with distance (3). We could have used cross-spectrum matrices as well. However, cross-spectral quantities estimated from short-windows of data are often very poor estimators and do not improve significantly the analysis. In fact, covariance matrices contain information about the sum of the contribution on each frequency of $\boldsymbol{x}(n)$'s spectrum, a feature that is often enough to track important changes to its statistics.

The results on sleep recordings are displayed in Figure 3. We did not include the "S1" sleep stage trials because they are

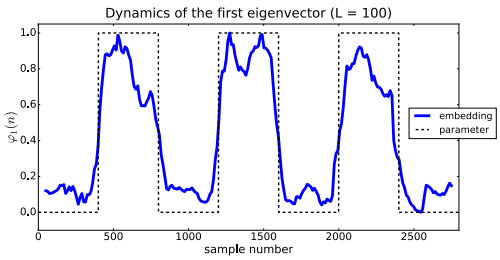


Fig. 2. Spectral embedding of the samples in a sliding-window ($L = 100$ and 90% overlap) running over the signals generated by (10). The blue curve corresponds to the first dimension of the spectral embedding of a window of data using distance (3) and the dashed line indicates the values of parameter $m(n)$.

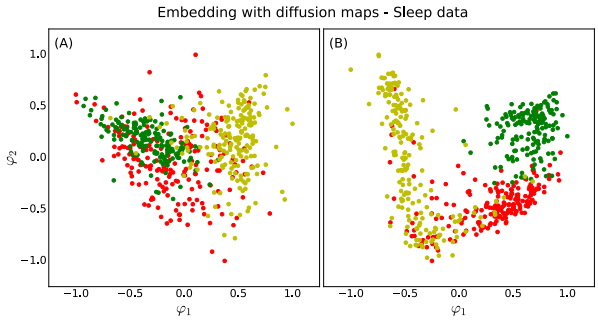


Fig. 3. Spectral embedding ($d = 2$) of the sleep data described in Sec. 2.4.2. Each point is a 30-sec clip and the colors indicate to which sleep stage it is associated (yellow for REM, red for S2, and green for S3). We used diffusion maps with two types of distance: (A) distance defined in (3) and (B) distance defined in (5).

associated to light sleep and can not be assumed G-WSS. We carried out the spectral embedding using distances (3) and (5) and observed a better separation of points with the latter. We link this to the traditional way of doing classification of sleep stages, which relies on the spectral content of the trials and indicates that their spectrum is an useful feature to classify them. These results are interesting because they show the possibility of extracting information from a dataset containing sleep recordings based only on an adequate notion of similarity between signals. The modified Rand score for a GMM clustering using distance (5) on the two first eigenvectors is 0.65 (in a scale from 0 to 1, where 1 is the best). We consider this a satisfying score because very few assumptions about the dataset were made and the classification was completely unsupervised.

In our last analysis, we wanted to track the statistics of the EEG of a subject when he was asked to toggle between eyes open or closed. We used a sliding-window containing 128 points (equivalent to 1 second) and applied the diffusion maps algorithm to embed the windows of data into a lower-

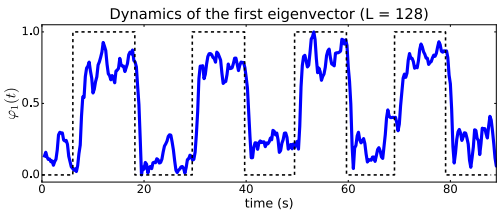


Fig. 4. First dimension of the spectral embedding of the samples in a sliding-window ($L = 128$ and 75% overlap) running over the EEG data described in Sec. 2.4.2. Each point of the blue curve corresponds to a different window of time. The dashed line indicates when the subject had the eyes open (0) and closed (1).

dimensional Euclidean space. The values of the first eigenvector of such spectral embedding are portrayed in Fig. 4, where we see a curve that clearly accompanies the changes in conditions of the subject's eyes.

The results presented above show that diffusion maps can be an useful tool for exploratory analysis of datasets containing time series. A remarkable feature of such approach is the lack of any model and very few assumptions about the time series. However, we should point out two drawbacks: there is no explicit mapping of the datapoints into the embedded space and it is often very difficult to give a physical interpretation to the axis of the spectral embedding.

4. CONCLUSION

We presented a new tool based on diffusion maps for analyzing multivariate time series. This method reduces drastically the number of dimensions necessary to analyse the dataset and allows for a concise description of its main features. We used two notions of distance between time series and discussed how to choose them according to the type of data and the kind of analysis one wants to perform. We illustrated our proposal on two examples with simulated signals and two with real data. Future investigations will focus on using our approach for improving classification pipelines of multivariate time series. We also intend to further investigate what information the trajectories in the embedded space can give and how they can be interpreted physiologically in the case of EEG time series.

5. ACKNOWLEDGMENT

The authors would like to thank Grégoire Cattan for helping to record the signals of the example in Figure 4.

6. REFERENCES

[1] Stephane Lafon, *Diffusion Maps and Geometric Harmonics*, Ph.D. thesis, Yale University, 2004.

[2] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 2007.

[3] Mikhail Belkin and Partha Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[4] Ronen Talmon, Stephane Mallat, Hitten Zaveri, and Ronald R. Coifman, "Manifold Learning for Latent Variable Inference in Dynamical Systems," *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3843–3856, 2015.

[5] M. B. Priestley, *Spectral Analysis and Time Series, Two-Volume Set, Volume 1-2: Volumes I and II*, Academic Press, 1983.

[6] Shun ichi Amari, *Information Geometry and Its Applications (Applied Mathematical Sciences)*, Springer, 2016.

[7] Marco Congedo, Alexandre Barachant, and Rajendra Bhatia, "Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, pp. 1–20, 2017.

[8] M. Congedo, P.L.C. Rodrigues, F. Bouchard, and A. Barachant, "A Closed-Form Unsupervised Geometry-Aware Dimensionality Reduction Method in the Riemannian Manifold of SPD Matrices," in *International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'17*, 2017.

[9] Mario Giovanni Terzano, Liborio Parrino, Adriano Sherieri, Ronald Chervin, Sudhansu Chokroverty, Christian Guilleminault, Max Hirshkowitz, Mark Mahowald, Harvey Moldofsky, Agostino Rosa, Robert Thomas, and Arthur Walters, "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Medicine*, vol. 2, no. 6, pp. 537–553, nov 2001.

[10] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, jun 2000.

[11] Lawrence Hubert and Phipps Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, dec 1985.