VISUALIZING HIGH DIMENSIONAL DYNAMICAL PROCESSES

Andrés F. Duque, Guy Wolf** and Kevin R. Moon**

*Utah State University, **Université de Montréal

Introduction

Manifold learning techniques have become of great interest when studying high dimensional data.
- Usually, the data have an extrinsic dimensionality that is artificially high, while its intrinsic structure is well-modeled as a low-dimensional manifold plus noise.
- Following the same line of reasoning, dynamical systems and time series can be regarded as processes governed by few underlying parameters, confined in a low-dimensional manifold.

Goal: Discover low-dimensional representations of high dimensional dynamical systems.

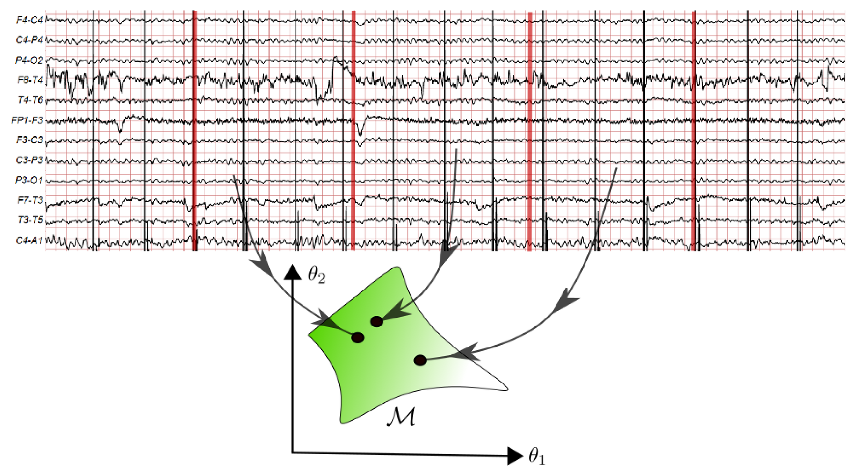

Figure 1: The process can be visualized in a low dimensional representation in the space of some underlying parameters $\boldsymbol{\theta}$ driving the system. The figure shows a 12-dimensional dynamical system. And the goal is to represent time windows of the data as points in a low-dimensional manifold.

Contributions:
- A novel manifold learning technique for dynamical systems called **DIG** (Dynamical information geometry).
- Incorporation of a novel group of distances in the context of diffusion operators.

Manifold Learning with Diffusion Operators

The use of diffusion operators in manifold learning was first introduced in Diffusion Maps [1]. Recently, more suited algorithms for visualization have been presented, such as PHATE [3]:

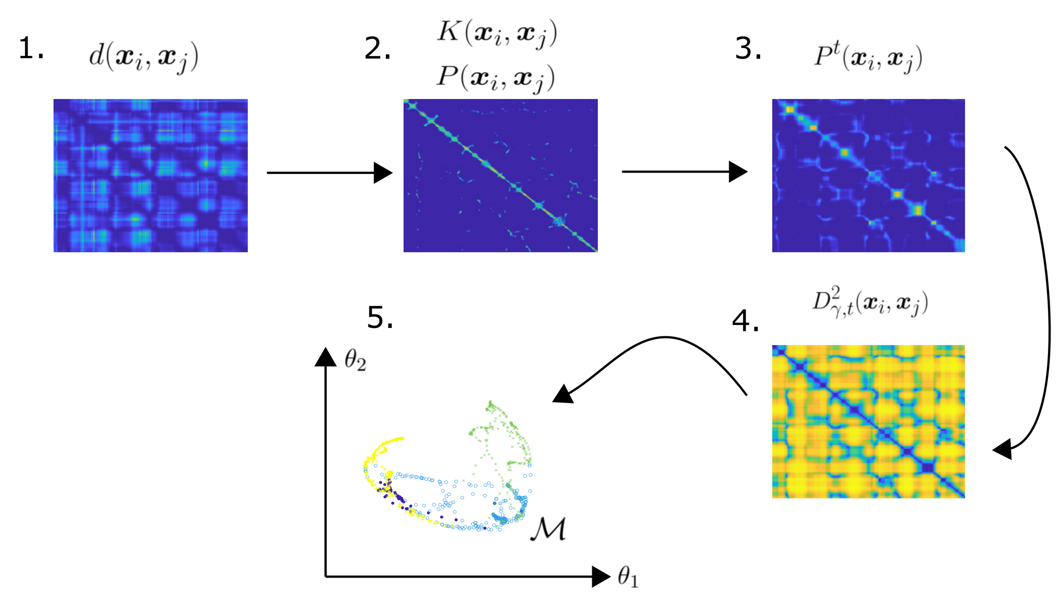

Figure 2: **PHATE steps.1.** Compute the distances between observations, typically the euclidean distance is employed in this step. **2.** Apply an adaptive kernel function to $d(x_i, x_j)$, and then row-normalize it to create a row stochastic matrix called the potential operator. **3.** Diffuse the potential operator t-steps forward. **4.**, Compute an information distance between the rows of P^t. Finally in **5**, apply metric MDS to $D^2_{\gamma,t}$.

- Diffusion Maps encapsulates the information in many dimensions.
- PHATE captures information in fewer dimensions $\Rightarrow$ better for visualization.

Diffusion with Dynamical Systems

In the context of dynamical systems we learn the local structure by constructing a matrix that encodes the local distances between time windows of data.

State-space formalism:

$$\boldsymbol{x}_t = \boldsymbol{y}_t(\boldsymbol{\theta}_t) + \boldsymbol{\xi}_t \tag{1}$$

$$d\theta^i_t = a^i(\theta^i_t)dt + dw^i_t, \quad i = 1, \ldots, d. \tag{2}$$

- $p(\boldsymbol{x}|\boldsymbol{\theta})$ is a linear transformation of $p(\boldsymbol{y}|\boldsymbol{\theta})$

New feature space, obtained by the histogram bins of the data within time windows of length L_1 centered at $\boldsymbol{x}_t$:

$$\boxed{\boldsymbol{x}_t \Rightarrow \boldsymbol{h}_t}$$

- The expected value of the histograms, e.g. $\mathbb{E}(h^j_t)$, is a linear transformation of $p(\boldsymbol{x}|\boldsymbol{\theta})$
- The Mahalanobis distance is invariant under linear transformations.

$\Rightarrow$ Distance (3) is noise resilient [4, 5]

$$d^2(\boldsymbol{x}_t, \boldsymbol{x}_s) = (\mathbb{E}(\boldsymbol{h}_t) - \mathbb{E}(\boldsymbol{h}_s))^T(\boldsymbol{C}_t)^{-1}(\mathbb{E}(\boldsymbol{h}_t) - \mathbb{E}(\boldsymbol{h}_s)), \tag{3}$$

Alternative distance:

Assumes that the data within time windows of length L_1 centered at x_t follows a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma_t)$.
- The geodesic distance between different time windows of data centered at $\boldsymbol{x}_t$ and $\boldsymbol{x}_s$ using the Fisher information as the Riemannian metric is as follows:

$$d^2(\boldsymbol{x}_t, \boldsymbol{x}_s) = \frac{1}{2}\sum_{i=1}^{N} \ln(\lambda_i), \text{ where } |\Sigma_t - \lambda_i \Sigma_s| = 0 \tag{4}$$

Information Distances

DIG extracts the information from the diffusion operator by embedding an information distance. We focus on a broad family of information distances that are parametrized by γ:

$$D^2_{\gamma,t}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} \sum_{k=1}^{N} \frac{(\log P^t_{ki} - \log P^t_{kj})^2}{\phi_0(k)}, & \gamma = 1 \\ \sum_{k=1}^{N} \frac{(P^t_{ki} - P^t_{kj})^2}{\phi_0(k)}, & \gamma = -1 \\ \sum_{k=1}^{N} \frac{2((P^t_{ki})^{\frac{1-\gamma}{2}} - (P^t_{kj})^{\frac{1-\gamma}{2}})^2}{(1-\gamma)\phi_0(k)}, & -1 < \gamma < 1. \end{cases} \tag{5}$$

Additionally, the rows of the diffusion matrix P can be interpreted as multinomial distributions. The geodesic distance between them using the Fisher information as the Riemannian metric is as follows:

$$\boxed{D(\boldsymbol{x}_i, \boldsymbol{x}_j) = 2\cos^{-1}\left(\sum_{k=1}^{N} \sqrt{P^t_{ki} P^t_{kj}}\right).} \tag{6}$$

After the information distances have been obtained, DIG applies metric multidimensional scaling (MDS) to the information distances to obtain a low-dimensional representation

Algorithm

Algorithm 1 The DIG algorithm

Input: Data matrix X, neighborhood size k, locality scale α, time windows length L_1 and L_2, number of bins Nb, information parameter γ, desired embedding dimension m (usually 2 or 3 for visualization)

Output: The DIG embedding Y_m

$d \leftarrow$ compute pairwise distance matrix from X using the mahalanobis distance (3)
$K_{k,\alpha} \leftarrow$ compute local affinity matrix from d and σ_k
$P \leftarrow$ normalize $K_{k,\alpha}$ to form a Markov transition matrix (diffusion operator)
$t \leftarrow$ compute time scale via Von Neumann Entropy [3]
Diffuse P for t time steps to obtain P^t
$D^\gamma_t \leftarrow$ compute the information distance matrix in eq. 5 from P^t for the given γ
$Y' \leftarrow$ apply classical MDS to D^γ_t
$Y_m \leftarrow$ apply metric MDS to D^γ_t with Y' as an initialization

Results in real data

We applied DIG to EEG data provided by [6, 2]. The data is labeled with one of six sleep categories according to R&K rules (REM, Awake, S-1, S-2, S-3, S-4). Due to the lack of observations in some stages, we group S-1 with S-2, and S-3 with S-4. We band-filtered the data between 8-40 Hz, and down-sampled it to 128Hz.

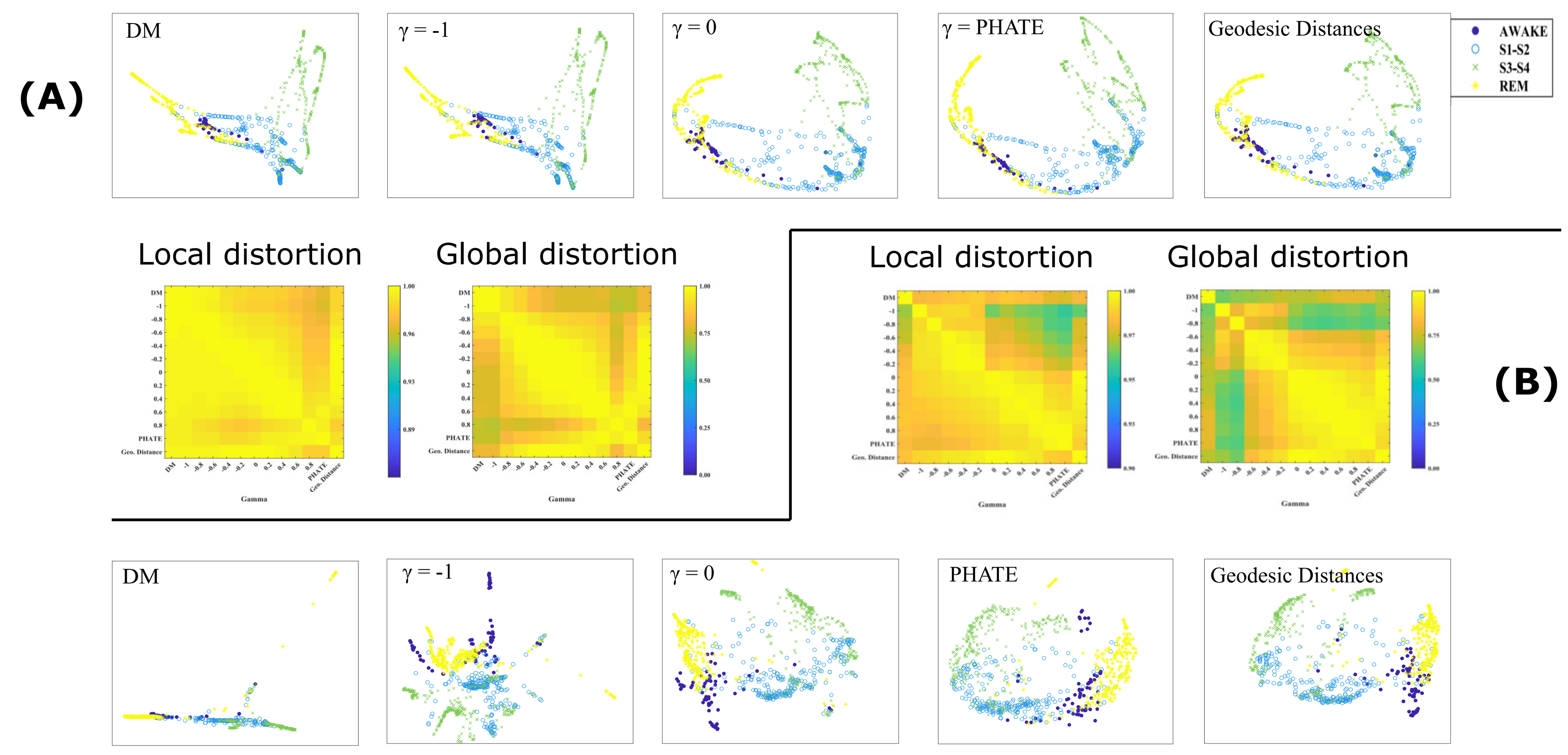

Figure 3: **(A)** Shows the embeddings obtained by the mahalanobis distance (3), for different values of γ. Additionally, we compare the relative local and global distortion of the embeddings, measured by the Trustworthiness and the Mantel test respectively. The same is replicated in **(B)** but for the gaussian information distance (4).

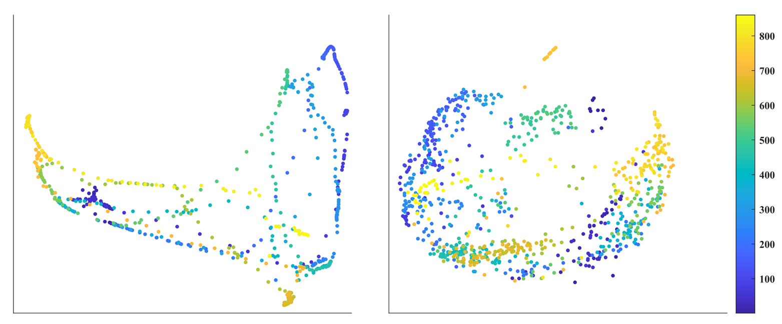

Figure 4: Visualization of EEG data colored by time steps using distance (3) at the left, and distance (4) at the right. Here we see how the left visualization presents a more denoised version, with clearer time-evolving transitions.

- In Figure 3 **(A)**, higher values of γ show the central structure of the embeddings more clearly defined than when using DM or lower values of γ.
- In Figure 3 **(B)**, the traditional DM tends to condense the structure together, and the use of the alternative γ values may reveal more details of the structure of the data. The most left embedding is a clear representation of such a situation, where DM does not show a suitable discrimination of the sleep stages. But when the value of gamma is increased, a more suitable representation is achieved.

Conclusions

- We derived a manifold learning tool called DIG for visualizing dynamical processes based on a diffusion framework. We addressed some of the shortcomings of the traditional diffusion maps approach for visualization.
- We presented experimental results where we were able to discover sleep dynamics using solely EEG recordings, as well as the time-varying progress of the processes.
- We presented a new group of distances in the context of diffusion operators.

References

[1] R.R. Coifman and S. Lafon. "Diffusion maps". In: *Applied and computational harmonic analysis* 21.1 (2006), pp. 5–30.

[2] A.L. Goldberger et al. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals". In: *Circulation* 101.23 (2000), e215–e220.

[3] K.R. Moon et al. "Visualizing transitions and structure for biological data exploration". In: *Nature Biotechnology (to appear)* (2019).

[4] R. Talmon and R.R. Coifman. "Empirical intrinsic geometry for nonlinear modeling and time series filtering". In: *Proceedings of the National Academy of Sciences* 110.31 (2013), pp. 12535–12540.

[5] R. Talmon and R.R. Coifman. "Intrinsic modeling of stochastic dynamical systems using empirical geometry". In: *Applied and Computational Harmonic Analysis* 39.1 (2015), pp. 138–160.

[6] M.G. Terzano et al. "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep". In: *Sleep medicine* 3.2 (2002), pp. 187–199.