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 modelled 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 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 [2]:

1. \( d(x, x_i) \)
2. \( P(x, x_i) \)
3. \( p^t(x, x_i) \)
4. \( h^t(x) \)
5. \( \phi(\gamma) \)

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 \( x_i(x, x_j) \), 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 \( D^t \).
- 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:

\[
\begin{align*}
& x_i(t) = \phi_i(T_i) + \xi \quad (1) \\
& df_i(t) = \phi_i'(T_i) dt + d\xi_i, \quad i = 1, \ldots, d \\
& d\phi_i(x) \text{ is a linear transformation of } \phi_i(t) \\
& \text{The expected value of the histograms, e.g. } \mathbb{E}[h_i(t)], \text{ is a linear transformation of } p^t(\theta) \\
& \text{The Mahalanobis distance is invariant under linear transformations.} \\
& \text{Distance } (3) \text{ is noise resilient [4, 5]} \\
& \forall \gamma, D^t(x_i, x_j) = \mathbb{E}[h_i(T_i)] - \mathbb{E}[h_j(T_j)] + \mathbb{E}[(C_i - h_i(T_i)) - (C_j - h_j(T_j))] \\
& \text{Alternative distance:} \\
& \text{Assumes that the data within time windows of length } L_t \text{ centered at } x_i \text{ follows a multivariate Gaussian distribution } N(\mu, \Sigma) \\
& \text{The geodesic distance between different time windows of data centered at } x_i \text{ and } x_j \text{, using the Fisher information as the Riemannian metric is as follows:} \\
& \forall \gamma, D^t(x_i, x_j) = \sum_{i=1}^{N} \lambda_i \log \lambda_i \gamma^2 \quad \text{where } |\lambda_i - \lambda_j| = 0 \\
\end{align*}
\]

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 \( \gamma \):

\[
D^t(x_i, x_j) = \begin{cases} 
\sum_{i=1}^{N} \frac{\lambda_i}{\gamma - 1} & \gamma > 1 \\
\sum_{i=1}^{N} \frac{\lambda_i}{\gamma + 1} & \gamma < 1
\end{cases}
\]

For \( \gamma = -1 \) \( \Rightarrow \) Geodesic Distances

For \( \gamma = 0 \) \( \Rightarrow \) PHATE Geodesic Distances

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

Algorithm

Input: Data matrix, neighborhood size, time windows length, \( t \), and number of bins \( N \). Information parameter \( \gamma \), desired embedding dimension \( m \) (usually 3 or 2 for visualization).
Output: The DIG embedding \( \gamma \).
1. \( \delta \text{ compute pairwise distance matrix from } \delta \text{ using the mahalanobis distance (3)} \)
2. \( \delta_i \text{ compute local affinity matrix from } \delta \text{ and } \theta_i \)
3. \( \delta_i \text{ normalize } \delta_i \text{ to form a Markov transition matrix (diffusion operator)} \)
4. \( \delta \text{ compute time scale via Von Neumann Entropy (2)} \)
5. \( \delta_i \text{ Diffuse } \delta_i \text{ by time steps to obtain } \delta_i \)
6. \( \delta_i \text{ compute the information distance matrix} \text{ in eq. 5 from } \delta_i \text{ for the given } \gamma \)
7. \( \delta_i \text{ apply classical MDS to } \delta_i \)
8. \( \delta_i \text{ apply metric MDS to } \delta_i \text{ with } \gamma \) 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 \( \gamma \). 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).

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