# Unsupervised Anomaly and Target Detection using Manifold Learning with Application to Deep Brain Stimulation (DBS)

Ido Cohen

# Unsupervised Anomaly and Target Detection using Manifold Learning with Application to Deep Brain Stimulation (DBS)

Research Thesis

As Partial Fulfillment of the Requirements for
the Degree Master of Science in Electrical Engineering

Ido Cohen

Submitted to the Senate of the Technion—Israel Institute of Technology
Iyar 5780 Haifa April 2020

# Acknowledgment

This research was carried out under the supervision of Prof. Ronen Talmon, in the Viterbi Faculty of Electrical Engineering. The generous financial help of the Technion is gratefully acknowledged.

# Contents


**1 Introduction** **17**
- 1.1 Motivation and Overview . . . 17
- 1.2 Thesis Structure . . . 18

**2 Scientific Background** **19**
- 2.1 Itô Process . . . 19
- 2.2 Diffusion Maps . . . 20
- 2.3 Nonlinear Independent Component Analysis . . . 21

**3 Proposed Method** **25**
- 3.1 Problem Formulation . . . 25
- 3.2 Extracting the System State Variables . . . 26
- 3.3 Proposed Algorithm . . . 31

**4 Method Illustration** **35**
- 4.1 Simulation . . . 35
- 4.2 Toy Mechanical System . . . 36

**5 Deep Brain Stimulation** **41**
- 5.1 Subthalamic Nucleus (STN) Detection . . . 41
- 5.2 Dorsolateral Oscillatory Region (DLOR) Detection . . . 43
- 5.3 Quantitative Detection Results . . . 45
- 5.4 Globus Pallidus (GP) Detection . . . 47

**6 Conclusion and Future work** **53**

# Abstract

Target and anomaly detection refers to the problem of finding unique patterns in data or patterns that do not conform to the expected behavior. Anomaly and target detection is considered challenging, especially since anomalies and targets can be expressed in various forms and anomalies are usually very rare. Most target and anomaly detection algorithms use prior knowledge such as predefined models and labeled data. This could lead to a significant bias, and the subsequent detection performance greatly depends on the quality of the prior knowledge.

To alleviate such a dependence and bias, we develop a data-driven unsupervised method based on manifold learning. We propose to define features which carry sufficient information that can be computed solely from the measurements. Then, we develop a variant of the Mahalanobis distance between these features and incorporate it into a specific manifold learning method, called diffusion maps. We analyze the proposed method using stochastic calculus. Particularly, we show that the modified Mahalanobis distance between the proposed features allows us to approximate the intrinsic variables that characterize the data, and by that facilitates an accurate target and anomaly detection.

We showcase our method on two applications. First, we apply it to observations of a simple mechanical system. This particular mechanical system was chosen since it has a known model, which serves as a definitive ground truth that can be used for validation. Using our method, we show the recovery of the main properties of the system solely from its observations in a data-driven manner. Indeed, the recovered properties coincide with the ground truth.

Second, we address a target detection task involving Deep Brain Stimulation (DBS). Typically in DBS, a surgery to implant a stimulating device is carried out. This device sends electrical signals to specific brain areas that are responsible for body movements. Once the device is implanted in an appropriate position, DBS can help in reducing the symptoms of tremor, slowness, stiffness, and walking problems caused by several neuronal diseases, such as Parkinson's disease, dystonia, or essential tremor. During the surgery, one important task is to detect the appropriate area for implanting the device. We focus on Parkinson's disease, for which the target region is the Sub-Thalamic Nucleus (STN) and a sub-region within the STN, called Dorso-Lateral Oscillatory Region (DLOR). An accurate detection of the STN and the DLOR is crucial for adequate clinical outcomes. Based on our method, we develop an unsupervised algorithm for the detection of the STN and the DLOR during a DBS surgery. We show that our algorithm attains detection results that outperform the gold standard. In addition, we show a proof of concept extension of the algorithm to the detection of the Globus Pallidus (GP) region that is of interest for treating dystonia.

# Notations

| | |
|---|---|
| $N$ | number of states in a dynamical system states |
| $M$ | number of observations from each state |
| $d_1$ | number of state variables |
| $d_2$ | number of noise variables |
| $d$ | total number of latent variables |
| $s$ | dimension of the observations |
| $\boldsymbol{\theta}_i(t)$ | state variables at time $t$ from the $i$th state |
| $\boldsymbol{\eta}_i(t)$ | noise variables at time $t$ from the $i$th state |
| $\boldsymbol{x}_i(t)$ | latent variables at time $t$ from the $i$th state |
| $\boldsymbol{y}_i(t_j)$ | the $j$th observation of the $i$th state |
| $\bar{\boldsymbol{\theta}}_{\boldsymbol{i}}$ | state variables baseline value at the $i$th state |
| $\bar{\boldsymbol{\beta}}_{\boldsymbol{i}}$ | noise variables baseline value at the $i$th state |
| $\bar{\boldsymbol{x}}_{\boldsymbol{i}}$ | latent variables baseline value at the $i$th state |
| $\boldsymbol{w}_{i,\theta}$ | $d_1$ uncorrelated standard Brownian motion |
| $\boldsymbol{w}_{i,\eta}$ | $d_2$ uncorrelated standard Brownian motion |
| $\delta t$ | time margin between consecutive observations |
| $\boldsymbol{z}_i$ | representation on all the measured data at the $i$th state |
| $f$ | a non linear measurement function |
| $C_i$ | the measurements covariance matrix of state $i$ |
| $x_i^k$ | the $k$th element of $\boldsymbol{x}_i$ |
| $C^{k_1 k_2}$ | the $(k_1, k_2)$ elements in the matrix C |
| $\|\cdot\|_M$ | modified Mahalanobis distance function |
| $\mathbf{J}$ | Jacobian matrix of $f$ |
| $\hat{\boldsymbol{z}}_i$ | estimator of state representation $\boldsymbol{z}_i$ |
| $\hat{\boldsymbol{C}}_i$ | estimator of the covariance matrix |
| $\Delta \boldsymbol{y}_i(t_j)$ | increment between 2 consecutive observations |
| $\tilde{\boldsymbol{C}}_i$ | approximation of the covariance matrix |
| $\boldsymbol{W}$ | affinity matrix between all system states |
| $\boldsymbol{K}$ | diffusion kernel |
| $\boldsymbol{K}_s$ | a smoothing diffusion kernel |
| $\boldsymbol{\Psi}_i$ | a global representation of the system in state $i$ |
| $\left\{\lambda_l^{(\cdot)}\right\}_{l=0}^{N-1}$ | set of eigenvalues of the diffusion kernel $K$ |
| $\left\{\psi_l^{(\cdot)}\right\}_{l=0}^{N-1}$ | set of eigenvectors of the diffusion kernel $K$ |
| $\boldsymbol{g}_i(t)$ | time series measurements of state $i$ before a pre processing stage |
| $\boldsymbol{d}_s$ | vector indicating the Estimated Distance from Target (EDT) |

# Abbreviations

| | |
|---|---|
| DBS | Deep Brain Stimulation |
| STN | Subthalamic Nucleus |
| DLOR | DorsoLateral Oscillatory Region |
| VMNR | Ventro Medial Non-oscillatory Region |
| EDT | Estimated Distance from Target |
| HMM | Hidden Markov Model |
| SNR | Signal to Noise Ratio |
| GP | Globus Pallidus |
| ODE | Ordinary Differential Equations |
| SDE | Stochastic Differential Equations |
| ICA | Independent Component Analysis |
| USVA | Unsupervised State Variables Approximation |

# List of Figures


4.1 An example of the evolution in time of the state variables in the simulation. 36
4.2 An example of the measurements of the simulated system. . . . . 37
4.3 An example of the obtained representation of the simulated system with respect to the ground truth. . . . . 38
4.4 A diagram of the mechanical system used as a toy problem . . . . . 39
4.5 An example of the mechanical system measurements . . . . . 40
4.6 The obtained representation of the measurements of the mechanical system. 40

5.1 An example of the input data to the STN detection algorithm. . . . . 42
5.2 STN detection – the 1D embedding obtained by the proposed method. . 43
5.3 Dorsolateral oscillatory region (DLOR) detection – the 2D embedding obtained by the proposed method. . . . . 45
5.4 STN detection illustration. . . . . 46
5.5 STN detection – performance comparison with the gold standard. . . . . 49
5.6 An example of the input data in the GP detection task. . . . . 50
5.7 GP detection – the 2D representation obtained by the proposed method. 50
5.8 GP detection illustration. . . . . 51

# 1 Introduction

## 1.1 Motivation and Overview

Target and anomaly detection refers to the problem of finding specific patterns of interest or abnormalities that do not conform to the expected or usual behaviour. This task is considered particularly challenging because targets may assume various forms, while anomalies are usually very rare. In addition, measurement systems typically introduce many sources of variability, where only few of them facilitate the distinction between routine and abnormal behaviour.

Most target and anomaly detection algorithms use prior knowledge on the data [1], such as predefined models and labels, in order to identify a specific or abnormal pattern. This could lead to bias models that greatly depend on the quality of the prior knowledge. To alleviate such a dependence and bias, we develop a data-driven unsupervised method that is able to reveal the sources of variability that characterize the intrinsic state of the system, and by that, facilitate an accurate target and anomaly detection based on the data.

Our approach to the problem makes use of manifold learning, which is a class of non-linear geometry-oriented dimensionality reduction methods, e.g., [2, 3, 4, 5]. Typically in manifold learning, a high dimensional data set assumed to lie on a low dimensional manifold is given. This class of methods attempt to reveal the intrinsic structure of the data set (the low dimensional manifold) by preserving distances within local neighborhoods. Manifold learning methods have been successfully applied to a broad range of applications, e.g., the discovery of the latent variables of dynamical systems [6, 7], image reconstruction [8], signal denoising [9], numerical simulation enhancement [10], fetal electrocardiogram analysis [11], sleep stage identification [12], and time series filtering [13], to name but a few.

In this thesis, we present a new target detection method based on Diffusion Maps [14], which is a pivotal manifold learning technique. Our method can be viewed as a follow up work of [15] and [16], where a new distance metric based on a variant of the Mahalanobis distance is incorporated into the diffusion maps framework. We show both theoretically and in experiments that the proposed method reveals the hidden source of variability, which characterizes the intrinsic state of the observed system, and by that, helping to detect target an anomalous patterns.

In addition, we address a particular target detection task involving Deep Brain Stimulation (DBS). Typically for DBS, a surgery to implant a stimulating device that sends electrical signals to brain areas that are responsible for body movements is carried out. Once the device is implanted in an appropriate position, DBS can help reducing the

symptoms of tremor, slowness, stiffness, and walking problems caused by several neuronal diseases, such as Parkinson's disease, dystonia, or essential tremor. During the surgery, one important task is to detect the appropriate position for implanting the device. The detection is based on micro electrode recordings along a pre-planned trajectory. The micro electrode recordings are typically intricate and one needs to extract the relevant information for the detection of the target regions. For this purpose, based on our target detection method, we implement purely unsupervised algorithms for finding specific regions, such as the subthalamic nucleus (STN) and a sub-region within the STN called DorsoLateral Oscillatory Region (DLOR), during a DBS surgery. We show that the performance of the proposed algorithm is comparable with the gold-standard in the detection of the STN. Importantly, the gold-standard today is based on a supervised Hidden Markov Model(HMM) algorithm, whereas the proposed algorithm is unsupervised, and therefore, is not biased toward a specific expert or depend on the existence of labels. In addition, we show that the proposed algorithm outperforms the supervised HMM algorithm in the detection of the DLOR. To illustrate the generality of the proposed algorithm, we present a proof-of-concept comprising the application of the proposed algorithm to the detection of the pallidal boundaries in the Globus Pallidus (GP) during DBS surgeries.

## 1.2 Thesis Structure

This thesis is organized as follows. In Chapter 2 we present the relevant scientific background for this work. In Chapter 3, we formulate the problem and present stochastic analysis, based on which, we devise a data-driven algorithm for target and anomaly detection. In Chapter 4, we demonstrate the algorithm on a simulation and on real recordings from a simple mechanical system. We show that the application of the algorithm enables to extract the system true model in an unsupervised manner without rigid prior knowledge, solely from measurements. In Chapter 5, we apply the proposed method to the problem of unsupervised detection of the brain areas during a DBS surgery. Finally, in Chapter 6, we conclude and discuss several directions for future work.

# 2 Scientific Background

## 2.1 Itô Process

Stochastic differential equations (SDEs) are often used for modeling the evolution in time of (possibly nonlinear) dynamical systems. Perhaps the most common class of such SDEs is based on the Itô process [17], which is defined next.

**Definition 1.** *An Itô process is a stochastic process $X(t)$ whose evolution in time is given by the following SDE:*

$$dX(t) = a(X(t), t)dt + b(X(t), t)dB(t) \tag{2.1}$$

*where $a$ and $b$ are real-valued functions on $\mathbb{R}^2$ and $B(t)$ is Brownian motion.*

Generally ,$a(\cdot, \cdot)$ and $b(\cdot, \cdot)$ are termed the *drift* and the *diffusion* of the process. An equivalent representation of $X(t)$ in (2.1) in an integral form is given by:

$$X(t) = \int_0^t a(X(s), s)ds + \int_0^t b(X(s), s)dB(s)$$

Since many dynamical systems are quite involved and the Itô process alone is insufficient to provide an informative modeling, real functions of the stochastic process $X(t)$ are considered. The resulting evolution in time is given by the Itô lemma, which could be viewed as the stochastic counterpart of the ordinary chain rule for calculating the derivative of a composite function.

**Theorem 2.** *Let $X(t)$ be an Itô process of the form 2.1 and let $f : \mathbb{R}^2 \to \mathbb{R}$ be a smooth function. Then, $Y(t) = f(X(t), t)$ is a stochastic process satisfiying the following SDE*

$$\begin{aligned} dY(t) = &\left(\frac{\partial f}{\partial t}(X(t), t) + a(X(t), t)\frac{\partial f}{\partial x}(X(t), t) + \frac{1}{2}b(X(t), t)\frac{\partial^2 f}{\partial x^2}(X(t), t)\right)dt \\ &+ b(X(t), t)\frac{\partial^2 f}{\partial x^2}(X(t), t)dB(t) \end{aligned} \tag{2.2}$$

In other words, $Y(t)$ is also an Itô process with a different drift and a different diffusion.

In the multidimensional case, where $f : \mathbb{R}^{d+1} \to \mathbb{R}$ is a smooth function and $Y(t)$ is a function of multiple Itô processes $X(t) = (X_1(t), ..., X_d(t))$ with a corresponding d-dimensional Brownian motion $B(t) = (B_1(t), ..., B_d(t))$, the Itô formula is extended and given by:

$$dY(t) = \frac{\partial f}{\partial t}(X(t), t)dt + \sum_{i=1}^{d}\frac{\partial f}{\partial X_i}(X(t), t)dX_i + \frac{1}{2}\sum_{i}^{d}\sum_{j}^{d}\frac{\partial^2 f}{\partial X_i \partial X_j}(X(t), t)dX_i dX_j \tag{2.3}$$

For further details on Itô calculus and SDEs, see [17].

## 2.2 Diffusion Maps

Manifold learning is a class of nonlinear geometry-oriented dimensionality reduction methods, e.g., [2, 3, 4, 5]. These methods aim at finding the intrinsic structure of high-dimensional data, which are assumed to lie on a low-dimensional manifold, by attempting to preserve distances within local neighborhoods. In this work, we focus on a specific manifold learning method, called Diffusion Maps [18]. In the remainder of this section, we briefly review its construction and describe its main properties.

Consider a set $\{\boldsymbol{x}_i\}_{i=1}^N$ of $N$ samples $\boldsymbol{x}_i \in \mathcal{M} \subset \mathbb{R}^s$, where $\mathcal{M}$ is a manifold embedded in an $s$-dimensional Euclidean space. The first step in the diffusion maps algorithm is to construct a weighted graph, where the nodes of the graph are the data samples $\{\boldsymbol{x}_i\}_{i=1}^N$, and the weights of the edges are determined by a kernel function with some distance metric. The typical kernel function is a Gaussian, giving rise to a graph affinity matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$, whose $(i,j)$th entry is given by

$$W_{i,j} = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_M^2}{\varepsilon}\right), \tag{2.4}$$

where $\varepsilon > 0$ is a parameter that determines the connectivity of the graph, and $\|\cdot\|_M$ is a distance metric, e.g., the Euclidean and the Mahalanobis distances [18, 15, 19, 20, 6].

Next, a diffusion operator $\mathbf{K} \in \mathbb{R}^{N\times N}$, is defined by normalizing the the affinity matrix $\mathbf{W}$ to be row-stochastic, that is a matrix whose sum of each row is equal to 1, in the following way:

$$Q_{i,i} = \left(\sum_{l=1}^{N} W_{i,l}\right)^{-1}; \mathbf{K} = \mathbf{Q}\mathbf{W}, \tag{2.5}$$

where $\mathbf{Q} \in \mathbb{R}^{N\times N}$ is a diagonal matrix, whose diagonal elements are the degrees of the nodes in the graph.

The matrix $\mathbf{K}$ can be interpreted as the transition matrix of a random walk defined on the graph, where $K_{i,j}$ represents the probability of transition from node $x_i$ to $x_j$ in a single random walk step. Accordingly, $\mathbf{K}^t$ for $t > 0$ can be viewed as the transition probability matrix of $t$ consecutive steps.

Finally, based on the spectral decomposition of $\mathbf{K}$, a nonlinear mapping that captures the manifold structure is defined by:

$$\boldsymbol{x}_i \mapsto [\lambda_1^t\psi_1(i), \lambda_2^t\psi_2(i), \ldots, \lambda_{N-1}^t\psi_{N-1}(i)] \triangleq \Psi_t(i), \tag{2.6}$$

where $\{\lambda_l, \boldsymbol{\psi}_l\}_{l=0}^{N-1}$ are the eigenvalues and the right eigenvectors of $\mathbf{K}$, indexed in descending order, and $\psi_l(i)$ denotes the $i$th element of $\boldsymbol{\psi}_l$. Note that $\lambda_0 = 1$ and that $\boldsymbol{\psi}_0$ is a trivial constant vector.

This nonlinear (diffusion maps) mapping facilitates the integration of the information from the entire set of points into the Euclidean distance between the maps of two individual samples. Concretely, let $d_t(i,j)$ be the diffusion distance [18] between $x_i$ and $x_j$,

which is defined by:

$$d_t(i,j) = \sqrt{\sum_{l=1,\ldots,N} \frac{\left((K^t)_{i,l} - (K^t)_{j,l}\right)^2}{\phi_0(l)}} \tag{2.7}$$

where $\phi_0(\cdot)$ is the stationary distribution of the random walk induced by $\mathbf{K}$. Indeed, by definition, the distance between $x_i$ and $x_j$ takes into account the transition probabilities (in $t$ steps) from these two sample to any other sample in the set. It was shown in [18] that the diffusion distance is equal to the Euclidean distance between the diffusion maps of the samples, i.e.:

$$d_t^2(i,j) = \|\Psi_t(i) - \Psi_t(j)\|^2 = \sum_{l=1}^{N-1} \lambda_l^{2t}(\psi_l(i) - \psi_l(j))^2. \tag{2.8}$$

In order to achieve a compact representation of the data, instead of using all of $N-1$ eigenvectors (excluding the trivial $\boldsymbol{\psi}_0$, the samples are mapped according to the $L$ most dominant components:

$$\boldsymbol{x}_i \mapsto [\lambda_1^t\psi_1(i), \lambda_2^t\psi_2(i), \ldots, \lambda_L^t\psi_L(i)], \tag{2.9}$$

In practice, $L$ is often determined by observing the decay of the eigenvalues of $\mathbf{K}$ and searching for a distinct "spectral gap", that is a significant gap between two consecutive eigenvalues, which could separate between 'signal' components and 'noise' components.

For more details on diffusion maps, see [18]. The diffusion maps algorithm appears in Algorithm 1.

## 2.3 Nonlinear Independent Component Analysis

Assume that the intrinsic variables that characterize the behaviour of a dynamical system lie on a low dimensional hidden manifold. Further assume that the accessible observations of the system are given through an unknown function that maps the intrinsic variables into some other high-dimensional manifold embedded in a Euclidean space.

In [15] the authors formulated the problem as a nonlinear independent component analysis (ICA) problem, namely the problem of finding the intrinsic variables that govern the dynamical system based on the observations, *assuming that these variables are independent*. More concretely, based on the assumption that the intrinsic hidden variables are generated by unknown stochastic and independent Itô processes, a method for recovering the hidden variables using diffusion maps is proposed.

Consider a model where the intrinsic variables are given by the following independent stochastic Itô processes:

$$dx_i = a_i(x_i)dt + b_i(x_i)dw_i, \quad i = 1, ..., n \tag{2.13}$$

where $a_i$ and $b_i$ are unknown drift and noise functions, and $w_i$ are independent Brownian motions. The n-dimensional process $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ is inaccessible. Instead, we

**Algorithm 1** Diffusion Maps

**Input**: High-dimensional samples $\{\boldsymbol{x}_i\}_{i=1}^N$, $\boldsymbol{x}_i \in \mathbb{R}^s$.
**Output**: $L$-dimensional representation $\{\Psi_t(i)\}_{i=1}^N$, where $\Psi_t(i) \in \mathbb{R}^L$.

1. Calculate the affinity matrix $\mathbf{W}$:

$$W_{i,j} = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_M^2}{\varepsilon}\right) \tag{2.10}$$

   where $\varepsilon <> 0$ is a scale parameter.

2. Compute the diffusion operator (transition matrix) $\mathbf{K}$:

$$Q_{i,i} = \left(\sum_{l=1}^{N} W_{i,l}\right)^{-1}; \mathbf{K} = \mathbf{QW}, \tag{2.11}$$

3. Calculate the spectral decomposition of $\mathbf{K}$ and obtain its eigenvalues $\{\lambda_l\}_{l=0}^{N-1}$ and right eigenvectors $\{\boldsymbol{\psi}_l\}_{l=0}^{N-1}$.

4. Define a new embedding for the samples:

$$\Psi_t(i) : \boldsymbol{x}_i \mapsto [\lambda_1^t\psi_1(i), \lambda_2^t\psi_2(i), \ldots, \lambda_L^t\psi_L(i)] \tag{2.12}$$

   where $t > 0$ is a parameter.

observe a non-linear mapping of $\boldsymbol{x}$, denoted by $\boldsymbol{y} = f(\boldsymbol{x})$, through the function $f : \mathbb{R}^n \to \mathbb{R}^m$ with $m \geq n$.

Using Itô lemma (2.2), it can be shown that in this case, the accessible $m \times m$ covariance matrix $\mathbf{C}$ of the observable process $\boldsymbol{y}$ is given by:

$$C_{jk} \coloneqq Cov(y_j, y_k) = \sum_{i=1}^{n} (b_i)^2 f_j^i f_k^i \tag{2.14}$$

where $f_j^i := \frac{\partial f^j}{\partial x^i}$. In matrix form, $\mathbf{C}$ is given by:

$$\mathbf{C} = \mathbf{JB}^2\mathbf{J}^T \tag{2.15}$$

where $\mathbf{J}$ is the $m \times n$ Jacobian matrix of $f$, and $\mathbf{B}$ is an $n \times n$ diagonal matrix with $B_{ii} = b_i$.

Utilizing (2.15) and a second-order Taylor expansion, the Euclidean distance between the hidden intrinsic variables can be approximated by

$$||\boldsymbol{\xi} - \boldsymbol{x}||^2 = d_m(\boldsymbol{y}, \boldsymbol{\eta}) + O(||\boldsymbol{\eta} - \boldsymbol{y}||^4) \tag{2.16}$$

where $\boldsymbol{y} = f(\boldsymbol{x})$ and $\boldsymbol{\eta} = f(\boldsymbol{\xi})$ are two observations of the hidden samples $\boldsymbol{x}$ and $\boldsymbol{\xi}$, respectively, and $d_m$ is a variant of the Mahalanonis distance, given by

$$d_m(\boldsymbol{y}, \boldsymbol{\eta}) = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{\eta})^T(\boldsymbol{C}_y^{-1} + \boldsymbol{C}_\eta^{-1})(\boldsymbol{y} - \boldsymbol{\eta}) \tag{2.17}$$

where $\boldsymbol{C}_y$ and $\boldsymbol{C}_\eta$ are the local covariance matrices at $\boldsymbol{y}$ and $\boldsymbol{\eta}$. The importance of (2.16) is that the right hand side can be computed from the accessible data, while the left hand side is the desired Euclidean distance between the hidden samples.

The approximation of the Euclidean distance between the hidden samples is utilized in an application of diffusion maps based on a diffusion kernel with the modified Mahalanobis distance (2.17) between the accessible observations. This application gives rise to a low dimensional representation of the intrinsic hidden variables.

This work by Singer and Coifman [15] has many extensions, e.g., [19, 21, 22, 23]. Of particular interest in the context of this work is [24], where the variant of the Mahalanobis distance (2.17) was used for the reduction of a multi-scale stochastic dynamical systems in a purely data-driven manner.

# 3 Proposed Method

## 3.1 Problem Formulation

Consider a system with $N$ different states, and denote the measurements of the system from each state by $\boldsymbol{y}_i(t)$ where $i = 1, \ldots, N$ denotes the index of the state and $t$ represents time. Suppose that our measurements have two sources of variability. The first source is governed by latent *state variables* $\boldsymbol{\theta}_i(t) \in \mathbb{R}^{d_1}$. These variables characterize the system state and their variation in time is a small perturbation of some baseline value $\bar{\boldsymbol{\theta}}_i \in \mathbb{R}^{d_1}$. Specifically, we assume that the evolution in time of the state variables can be described by the following Itô process:

$$d\boldsymbol{\theta}_i(t) = -\nabla \boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_i(t))dt + \boldsymbol{I}_{d_1 \times d_1} d\boldsymbol{w}_{i,\theta}(t) \tag{3.1}$$

where the process drift is the gradient of the quadratic potential function $\boldsymbol{U}_\theta(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_i)^T(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_i)$ centered at the baseline value $\bar{\boldsymbol{\theta}}_i$, $\boldsymbol{w}_{i,\theta}(t)$ is a vector of $d_1$ independent Brownian motions and $\boldsymbol{I}_{d_1 \times d_1}$ is a $d_1 \times d_1$ identity matrix.

The second source of variability is considered to be noise and represented by the latent variables $\boldsymbol{\eta}_i(t) \in \mathbb{R}^{d_2}$. Suppose that the noise variables are characterized by high variability in time compared to the variability of the state variables. Formally, the evolution in time of the noise variables can be described by the following Itô process:

$$d\boldsymbol{\eta}_i(t) = -\nabla \boldsymbol{U}_{\boldsymbol{\eta}}(\boldsymbol{\eta_i}(t))dt + \frac{1}{\epsilon}\boldsymbol{I}_{d_2 \times d_2} d\boldsymbol{w}_{i,\eta}(t) \tag{3.2}$$

where the drift is the gradient of a quadratic potential function given by $\boldsymbol{U}_{\boldsymbol{\eta}}(\boldsymbol{\eta}) = \frac{1}{2}(\boldsymbol{\eta} - \bar{\boldsymbol{\beta}}_i)^T(\boldsymbol{\eta} - \bar{\boldsymbol{\beta}}_i)$, $\bar{\boldsymbol{\beta}}_i$ is an unknown baseline constant, $\boldsymbol{I}_{d_2 \times d_2}$ is a $d_2 \times d_2$ identity matrix, $\boldsymbol{w}_{i,\eta}(t)$ is a vector of $d_2$ independent Brownian motions, and $0 < \epsilon << 1$. Note that the variance of the diffusion term of the noise variables in (3.2) is larger than the variance of the diffusion term of the state variables in (3.1) by a factor of $1/\epsilon^2$. We assume that the two Ito processes are uncorrelated. Similar models are often used to describe a multiscale stochastic dynamical systems [16].

The measurements are given by $\boldsymbol{y}_i(t) = f(\boldsymbol{\theta}_i(t), \boldsymbol{\eta}_i(t))$, where $f : \mathbb{R}^d \to \mathbb{R}^s$ is some, possibly nonlinear, function and $d = d_1 + d_2$. For notational convenience, we denote $\boldsymbol{x}_i(t) = (\boldsymbol{\theta}_i(t), \boldsymbol{\eta}_i(t))$, and accordingly, we recast (3.1) and (3.2) as the following Itô process in $d$ dimensions:

$$d\boldsymbol{x}_i(t) = -\nabla \boldsymbol{U}(\boldsymbol{x}_i(t))dt + \Lambda(d\boldsymbol{w}_i(t)) \tag{3.3}$$

where $\boldsymbol{U}(\boldsymbol{x})$ is a quadratic potential function centered at $\bar{\boldsymbol{x}}_i = \begin{bmatrix} \bar{\boldsymbol{\theta}}_i \\ \bar{\boldsymbol{\beta}}_i \end{bmatrix}$, $\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{I}_{d_1 \times d_1} & 0 \\ 0 & \frac{1}{\epsilon}\boldsymbol{I}_{d_2 \times d_2} \end{bmatrix}$ and $\boldsymbol{w}_i(t) = \begin{bmatrix} \boldsymbol{w}_{i,\theta}(t) \\ \boldsymbol{w}_{i,\eta}(t) \end{bmatrix}$.

We sample the system $M$ consecutive times from each state. Let $\boldsymbol{y}_i(t_j) \in \mathbb{R}^s$ denote the $j$th measurement of the system at the $i$th state, where $j = \{0, 1, 2, ...., M-1\}$ and $\delta t$ is the time margin between two consecutive measurements such that $t_j = \delta t \cdot j$ .

Our goal is to decouple the two sources of variability given the measurements $\boldsymbol{y}_i(t_j)$ without prior knowledge on the system variables, and to build a parametrization of the system state, which is consistent with the state variables. Since the state variables constitute an accurate proxy of the true state of the system, the ability to extract them facilitates the identification of particular desired system regimes and anomalous states.

In order to accomplish this goal, we devise a pairwise distance between system states that satisfies:

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) \approx \alpha ||\bar{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}}_l||^2 \tag{3.4}$$

where $\boldsymbol{z}_i$ is some representation of the measured data $\{\boldsymbol{y}_i(t_j)\}_{j=1}^M$ at the $i$th state and $\alpha$ is some constant.

The specification of the above problem formulation in the context of STN detection during DBS surgery is as follows. We measure from $N$ depths along the pre-planned trajectory and from each depth we acquire $M$ measurements, denoted by $\boldsymbol{y}_i(t_j)$, where $i$ is now the index of a specific depth. We assume that the measurements are driven by two sources of variability. The first source of variability is represented by the state variables $\boldsymbol{\theta}_i(t_j)$, which are some unknown hidden variables that characterize the STN region. The second source of variability is represented by the noise variables $\boldsymbol{\eta}_i(t_j)$. We do not have a direct access to the state variables depending on the region nor to the noise variables, and we measure them through some unknown possibly nonlinear function $f$ of $\boldsymbol{x}_i(t_j) = (\boldsymbol{\theta}_i(t_j), \boldsymbol{\eta}_i(t_j))$. From the measurements, we aim to find a parametrization of the system state, which will allow us to identify in a purely unsupervised manner the STN region.

## 3.2 Extracting the System State Variables

As stated above, our goal is to find a pairwise distance between system states that reveal the relation between the state variables based on the measurements. Our approach consists of two stages. First, we define features, which carry sufficient information about the state variables and can be computed solely from the measurements. We propose to use two features representing the measurements from each state:

$$\boldsymbol{z}_i = \mathbb{E}(\boldsymbol{y}_i(t_1)) \tag{3.5}$$

$$\boldsymbol{C}_i = \mathrm{Cov}(\boldsymbol{y}_i(t_1)) \tag{3.6}$$

namely, the expected value and the covariance matrix of the measurements of the Itô process at a specific state.

Second, we define an appropriate metric between these features, enabling us to reveal the state variables. Particularly, we propose to use a modified version of the Mahalanobis distance[15]:

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) = \frac{1}{2}(\boldsymbol{z}_i - \boldsymbol{z}_l)^T(\boldsymbol{C}_i^{-1} + \boldsymbol{C}_l^{-1})(\boldsymbol{z}_i - \boldsymbol{z}_l) \tag{3.7}$$

and show in the sequel that it indeed allows us to uncover the relations between the state variables.

**Direct Access.** First, for simplifying the exposition, we consider the case in which we have a direct access to the system variables $\boldsymbol{x}_i(t) = (\boldsymbol{\theta}_i(t), \boldsymbol{\eta}_i(t)) \in \mathbb{R}^d$. In this section we will demonstrate how the proposed features and metric in this case achieve our original goal by proving the following proposition.

**Proposition 3.** *Given $\boldsymbol{x}_i(t) = (\boldsymbol{\theta}_i(t), \boldsymbol{\eta}_i(t)) \in \mathbb{R}^d$ , the modified Mahalanobis distance in (3.7) between the features $\boldsymbol{z}_i = \mathbb{E}[\boldsymbol{x}_i(t_1)]$ using the covariance $\boldsymbol{C}_i = \mathrm{Cov}(\boldsymbol{x}_i(t_1))$ can be written in terms of the Euclidean distance between the underlying state variables as follows:*

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) = \frac{1}{\delta t}[||\bar{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}}_l||^2 + O(\epsilon)] \tag{3.8}$$

The proof of the proposition above relies on the following results.

**Lemma 4.** *The expected value of the Itô process $\boldsymbol{x}_i(t)$ in (3.3) is $E[\boldsymbol{x}_i(t)] = \bar{\boldsymbol{x}}_i$*

*Proof of Lemma 4.* The measurements evolution in time can be represented explicitly by:

$$d\boldsymbol{x}_i(t) = (\bar{\boldsymbol{x}}_i - \boldsymbol{x}_i(t))dt + \boldsymbol{\Lambda} d\boldsymbol{w}_i(t) \tag{3.9}$$

The evolution of the $k$th element of $\boldsymbol{x}_i(t)$ is given by

$$dx_i^k = (\bar{x}_i^k - x_i^k(t))dt + \sigma^k dw_i^k(t) \quad , \quad k = 1, ..., d \tag{3.10}$$

where $\sigma^k$ is the $k$-th element in the diagonal of $\boldsymbol{\Lambda}$. Taking the expected value of (3.10) yields:

$$\begin{aligned}\mathbb{E}(x_i^k)(t) &= x_i^k(0) + \mathbb{E}\int_0^t (\bar{x}_i^k - x_i^k(s))ds \\ &= x_i^k(0) + t\bar{x}_i^k - \int_0^t \mathbb{E}(x_i^k(s))ds\end{aligned}$$

and the solution of this ODE is:

$$\mathbb{E}(x_i^k)(t) = \bar{x}_i^k$$

In vector form, we have $\mathbb{E}(\boldsymbol{x}_i)(t) = \bar{\boldsymbol{x}}_i$. □

**Lemma 5.** *The covariance matrix of the Itô process* $\boldsymbol{x}_i(t)$ *in* (3.3) *at time* $t_1 = \delta t$ *is given by* $\mathrm{Cov}(\boldsymbol{x}_i(\delta t)) = \delta t \boldsymbol{\Lambda}^2$

*Proof of Lemma 5.* By definition, the $(k_1, k_2)$-th element of the covariance matrix is

$$\begin{aligned}\mathrm{Cov}(x_i(\delta t))^{k_1 k_2} &= \mathrm{Cov}(x_i^{k_1}(\delta t), x_i^{k_2}(\delta t)) \\ &= \mathbb{E}[x_i^{k_1}(\delta t) x_i^{k_2}(\delta t)] - \mathbb{E}[x_i^{k_1}(\delta t)]\mathbb{E}[x_i^{k_2}(\delta t)] \\ &= \delta t (\Lambda^{k_1 k_2})^2\end{aligned}$$

where $\Lambda^{k_1 k_2}$ is the $(k_1, k_2)$-th element of the matrix $\boldsymbol{\Lambda}$. In matrix form, we have:

$$\mathrm{Cov}(\boldsymbol{x}_i(\delta t)) = \delta t \boldsymbol{\Lambda}^2$$

□

*Proof of Proposition 3.* According to Lemma 4 and Lemma 5, the features in (3.5) and (3.6) are: $(\boldsymbol{z}_i, \boldsymbol{C}_i) = (\bar{\boldsymbol{x}}_i, \delta t \boldsymbol{\Lambda}_i^2)$. By using the inverse of the covariance matrix:

$$\boldsymbol{C}_i^{-1} = \frac{1}{\delta t}\begin{bmatrix} \boldsymbol{I} & 0 \\ 0 & \epsilon^2 \boldsymbol{I} \end{bmatrix}$$

and by substituting $(\boldsymbol{z}_i, \boldsymbol{C}_i)$ into (3.7) we obtain that:

$$\begin{aligned} d(\boldsymbol{z}_i, \boldsymbol{z}_l) &= \frac{1}{2}(\boldsymbol{z}_i - \boldsymbol{z}_l)^T (\boldsymbol{C}_i^{-1} + \boldsymbol{C}_l^{-1})(\boldsymbol{z}_i - \boldsymbol{z}_l) \\ &= \frac{1}{\delta t}\sum_{k=1}^{d} e^k ((\bar{x}_i)^k - (\bar{x}_l)^k)^2 \end{aligned}$$

where $e^k = 1$ if $k \in \{1, .., d_1\}$ (denoting the indices of the state variables), and $e^k = \epsilon$ if $k \in \{d_1 + 1, .., d_1 + d_2\}$ (denoting the indices of the noise variable). Further derivation yields:

$$\begin{aligned} d(\boldsymbol{z}_i, \boldsymbol{z}_l) &= \frac{1}{\delta t}\sum_{k=1}^{d} e^k ((\bar{x}_i)^k - (\bar{x}_l)^k)^2 \\ &= \frac{1}{\delta t}\left[\sum_{k=1}^{d_1}((\theta_i)^k - (\theta_l)^k)^2 + \sum_{k=d_1+1}^{d} \epsilon((\eta_i)^k - (\eta_l)^k)^2\right] \\ &= \frac{1}{\delta t}\left[||\boldsymbol{\theta}_i - \boldsymbol{\theta}_l||^2 + \epsilon||\boldsymbol{\beta}_i - \boldsymbol{\beta}_l||^2\right] \\ &= \frac{1}{\delta t}\left[||\boldsymbol{\theta}_i - \boldsymbol{\theta}_l||^2 + O(\epsilon)\right] \end{aligned}$$

□

For convenience, we denote the norm associated with the modified Mahalanobis distance as:

$$||\boldsymbol{z}_i - \boldsymbol{z}_l||_M = \sum_{k=1}^{d} e_k((z_i)^k - (z_l)^k)^2$$

Proposition 3 implies that by assuming direct access to the state and noise variables, the modified Mahalanobis distance with the proposed features satisfies our goal specified in (3.4).

**Non-Linear Measurements.** Now, we consider a more general case where the observations are a function of $\boldsymbol{x}$, i.e., $\boldsymbol{y}_i(t) = f(\boldsymbol{x}_i(t))$, where $f : \mathbb{R}^d \to \mathbb{R}^s$ is some smooth function. In this section we will show that under a certain assumption on the function $f$ and a small modification of the feature $\boldsymbol{z}_i$, the modified Mahalanobis distance (3.7) with the proposed features achieves the desired goal.

**Assumption 6.** *For any two realizations $\boldsymbol{x}_i$ and $\boldsymbol{x}_l$ of state and noise variables, we have:*

$$\frac{(f^k_{p_1p_2}(x_i) - f^k_{p_1p_2}(x_l))^2}{f^k_{p_1}(x_i)f^k_{p_2}(x_i) + f^k_{p_1}(x_l)f^k_{p_2}(x_l)} << 1, \qquad (3.11)$$
$$1 \le k \le s, \qquad 1 \le p_1, p_2 \le d,$$

*where the subscripts correspond to partial derivatives, i.e., $f^k_p = \frac{\partial f^k}{\partial x^p}$ and $f^k_{p_1p_2} = \frac{\partial^2 f^k}{\partial x^{p_1}\partial x^{p_2}}$, and recall that superscripts correspond to specific elements in a vector, i.e., $f^k : \mathbb{R}^d \to \mathbb{R}$ is the $k$-th element of $f$.*

**Proposition 7.** *Given observations $\boldsymbol{y}_i(t) = f(\boldsymbol{x}_i(t))$, where $f : \mathbb{R}^d \to \mathbb{R}^s$ is a smooth function satisfying Assumption 6, the modified Mahalanobis distance with the features*

$$\boldsymbol{z}_i = \mathbb{E}\left[\frac{1}{\delta t}(\boldsymbol{y}_i(t_1) - \boldsymbol{y}_i(0)) + \boldsymbol{y}_i(0)|\boldsymbol{x}_i(0)\right] \qquad (3.12)$$
$$\boldsymbol{C}_i = Cov(\boldsymbol{y}_i(t_1))$$

*can be expressed as:*

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) = ||\bar{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}}_l||^2 + O(||\boldsymbol{y}_i - \boldsymbol{y}_l||^4) + O(\epsilon) \qquad (3.13)$$

The proof of the proposition above relies on the following results.

**Lemma 8.** *The conditional expected value of $\boldsymbol{y}_i(\delta t)$ satisfies the following relation:*

$$\mathbb{E}\left[\frac{1}{\delta t}(\boldsymbol{y}_i(\delta t) - \boldsymbol{y}_i(0)) + \boldsymbol{y}_i(0)|\boldsymbol{x}_i(0)\right] = \bar{\boldsymbol{y}}_i + \boldsymbol{\phi}_i + O(||\bar{\boldsymbol{y}}_i - \boldsymbol{y}_i(0)||^2)$$

*where* $\bar{\boldsymbol{y}}_i = \begin{bmatrix} \bar{y_i}^1 \\ .. \\ \bar{y_i}^d \end{bmatrix}$ *and* $\boldsymbol{\phi}_i = \begin{bmatrix} \frac{1}{2}\sum_{p=1}^{d}(\sigma^k)^2 f_{pp}^1(\boldsymbol{x}_i(0)) \\ .. \\ \frac{1}{2}\sum_{k=1}^{d}(\sigma^k)^2 f_{pp}^d(\boldsymbol{x}_i(0)) \end{bmatrix}$.

*Proof of Lemma 8.* The process $\boldsymbol{y}_i(t)$ is given by the Itô Lemma as:

$$dy_i^k = [(\bar{\boldsymbol{x}}_i - \boldsymbol{x}_i(t))^T \nabla f^k(\boldsymbol{x}_i(t)) + \frac{1}{2}\sum_{p=1}^{d}(\sigma^k)^2 f_{pp}^k(\boldsymbol{x}_i(t))]dt + (\nabla f^k(\boldsymbol{x}_i(t)))^T \boldsymbol{\Lambda} d\boldsymbol{W}_i(t)$$

We note that the Itô process $\boldsymbol{x}_i(t)$ is defined as a diffusion process, and by [25] the expected value of $\boldsymbol{y}_i(t) = f(\boldsymbol{x}_i(t))$ is given by:

$$\mathbb{E}\left[\frac{y_i^k(\delta t) - y_i^k(0)|x_i(0)}{\delta t}\right] = (\bar{\boldsymbol{x}}_i - \boldsymbol{x}_i(0))^T \nabla f^k(\boldsymbol{x}_i(0)) + \frac{1}{2}\sum_{p=1}^{d}(\sigma^k)^2 f_{pp}^k(\boldsymbol{x}_i(0))$$

Using a first order Taylor expansion of $f^k(\boldsymbol{x}_i)$ around $\boldsymbol{x}_i(0)$ yields:

$$\mathbb{E}\left[\frac{y_i^k(\delta t) - y_i^k(0)|\boldsymbol{x}_i(0)}{\delta t}\right] + y_i(0)^k = \bar{y_i}^k + \frac{1}{2}\sum_{p=1}^{d}(\sigma^k)^2 f_{pp}^k(\boldsymbol{x}_i(0))$$

The proof is concluded by writing the above in matrix form:

$$\mathbb{E}\left[\frac{\boldsymbol{y}_i(\delta t) - \boldsymbol{y}_i(0)|\boldsymbol{x}_i(0)}{\delta t}\right] + \boldsymbol{y}_i(0) = \bar{\boldsymbol{y}}_i + \boldsymbol{\phi}_i + O(||\bar{\boldsymbol{y}}_i - \boldsymbol{y}_i(0)||^2)$$

□

**Lemma 9.** *The covariance of* $\boldsymbol{y}_i(t)$ *at* $\delta t$ *is given by:*

$$\boldsymbol{C} = \mathrm{Cov}(\boldsymbol{y}_i(\delta t)) = \boldsymbol{J}\boldsymbol{\Lambda}^2\boldsymbol{J}^T$$

*where* $\boldsymbol{J}$ *is the* $s \times d$ *Jacobian matrix of of* $\boldsymbol{y} = f(\boldsymbol{x})$.

The proof of Lemma 9 appears in [15].

*Proof of Proposition 7.* According to Lemma 8 and Lemma 9, the modified features (3.12) are: $(\boldsymbol{z}_i, \boldsymbol{C}_i) = (\bar{\boldsymbol{y}}_i + \boldsymbol{\phi}_i, \boldsymbol{J}\boldsymbol{\Lambda}^2\boldsymbol{J}^T)$.
The proposed distance with the modified features is given by:

$$\begin{aligned} d(\boldsymbol{z}_i, \boldsymbol{z}_l) &= \frac{1}{2}(\boldsymbol{z}_i - \boldsymbol{z}_l)^T(\boldsymbol{C}_i^{-1} + \boldsymbol{C}_l^{-1})(\boldsymbol{z}_i - \boldsymbol{z}_l) \\ &= \frac{1}{2}(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}}_l)^T((\boldsymbol{J}\boldsymbol{\Lambda}^2\boldsymbol{J}^T)^{-1} + (\boldsymbol{J}\boldsymbol{\Lambda}^2\boldsymbol{J}^T)^{-1})(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}}_l) \\ &+ \frac{1}{2}(\phi_i - \phi_l)^T((\boldsymbol{J}\boldsymbol{\Lambda}^2\boldsymbol{J}^T)^{-1} + (\boldsymbol{J}\boldsymbol{\Lambda}^2\boldsymbol{J}^T)^{-1})(\boldsymbol{\phi}_i - \boldsymbol{\phi}_l) \end{aligned} \tag{3.14}$$

According to [15], the first term in the right hand size of the equation is equal to:

$$\frac{1}{2}(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}}_l)^T((\boldsymbol{J}\boldsymbol{\Lambda}^2\boldsymbol{J}^T)^{-1} + (\boldsymbol{J}\boldsymbol{\Lambda}^2\boldsymbol{J}^T)^{-1})(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}}_l) = ||\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_l||_M + O(||\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}}_l||^4)$$

Applying Assumption 6 on the function $f$ to the second term in the right hand side of the equation is of order $O(\epsilon)$, where $\epsilon \ll 1$.

Combining the above, we obtain:

$$\begin{aligned} d(\boldsymbol{z}_i, \boldsymbol{z}_l) &= ||\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_l||_M + O(||\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}}_l||^4) + O(\epsilon) \\ &= ||\bar{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}}_l||^2 + O(||\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}}_l||^4) + O(\epsilon) \end{aligned}$$

□

We conclude that in both scenarios (with direct access or through some unknown observation function), the modified Mahalanobis distance between the proposed features (in (3.5) ,(3.6) and (3.12)) reveals the distance between the state variables and attenuates the contribution of the noise variables.

**Remark 10.** *In the specific case, where $f$ is the identity function, i.e $f(x) = x$ , a different feature $z_i$ is computed depending on the scenario (either (3.5) or (3.12)). Indeed, using the additional prior information (having a direct access to the state and noise variables) leads to a better approximation as evident by Proposition 3 and Proposition 7. We will later show that in practice the two definitions of $z_i$ coincide.*

## 3.3 Proposed Algorithm

We propose an unsupervised algorithm that is able to reveal the relations between the system intrinsic variables without any prior knowledge by relying on the analysis presented in Section 3.2. Broadly, the proposed algorithm consists of two main stages. In the first stage, we present estimators for the features defined in 3.12 and 3.6 from the measurements at hand. In the second stage, we apply a manifold learning method, diffusion maps, that constructs a global representation of the hidden state variables based on the estimated features and their pairwise distances.

**Features Estimators.** The proposed features estimators are based on two assumptions. First, we assume that that number of measurements at each state $M$ is large. Second, we assume that the measured signal $\boldsymbol{y}_i(t_j)$ is stationary with respect to $t_j$ in a fixed state $i$.

For $\boldsymbol{z}_i$ we propose the following estimator:

$$\hat{\boldsymbol{z}}_i = \frac{1}{M}\sum_{j=1}^{M} \boldsymbol{y}_i(t_j) \tag{3.15}$$

where the relation between the estimator and the desired feature can be expressed as:

$$\boldsymbol{z}_i = \hat{\boldsymbol{z}}_i + O\left(\frac{1}{\delta t M}\right) \tag{3.16}$$

Recall that $M$ is assumed to be large and therefore $\hat{\boldsymbol{z}}_i$ can be considered as a good approximation of $\boldsymbol{z}_i$. The relation in (3.16) stems from the following derivation. Since we assume that $M$ is large, then by the Law of Large Numbers, the empirical mean converges to the expected value, and so the desired feature can be recast as:

$$\begin{aligned}
\boldsymbol{z}_i &= \mathbb{E}\left[\frac{1}{\delta t}(\boldsymbol{y}_i(\delta t) - \boldsymbol{y}_i(0)) + \boldsymbol{y}_i(0)|\boldsymbol{x}_i(0)\right] \\
&= \frac{1}{M}\sum_{j=1}^{M}\left[\frac{\boldsymbol{y}_i(t_j) - \boldsymbol{y}_i(t_{j-1})}{\delta t} + \boldsymbol{y}_i(t_{j-1})\right] \\
&= \frac{1}{M}\sum_{j=1}^{M}\left[\frac{\boldsymbol{y}_i(t_j) - \boldsymbol{y}_i(t_{j-1})}{\delta t}\right] + \frac{1}{M}\sum_{j=1}^{M}\boldsymbol{y}_i(t_{j-1}) \\
&= \frac{1}{M}\sum_{j=1}^{M}\left[\frac{\boldsymbol{y}_i(t_j) - \boldsymbol{y}_i(t_{j-1})}{\delta t}\right] + \hat{\boldsymbol{z}}_i
\end{aligned} \tag{3.17}$$

The first term in (3.17) is a telescopic sum; after cancellation, this term is equal to $O(\frac{1}{\delta t M})$, leading to (3.16). We note that in practice the estimate of $\hat{\boldsymbol{z}}_i$ in the nonlinear case (3.12) coincides with the feature proposed for the linear case (3.5).

Denote the increments between consecutive measurements by $\Delta\boldsymbol{y}_i(t_j) = \boldsymbol{y}_i(t_j) - \boldsymbol{y}_i(t_{j-1})$. The estimator of $\boldsymbol{C}_i$ is the empirical covariance of the increments series:

$$\hat{\boldsymbol{C}}_i = \frac{1}{M-1}\sum_{j=1}^{M-1}[\Delta\boldsymbol{y}_i(t_j) - \hat{\Delta y_i})][\Delta\boldsymbol{y}_i(t_j) - \hat{\Delta y_i})]^T \tag{3.18}$$

where the $\hat{\Delta y_i}$ is the empirical mean of the increments, given by

$$\hat{\Delta\boldsymbol{y}_i} = \frac{1}{M}\sum_{j=1}^{M}\Delta\boldsymbol{y}_i(t_j)$$

The relation between the proposed estimator and the desired feature is:

$$\boldsymbol{C}_i = \hat{\boldsymbol{C}}_i + O(\delta t) \tag{3.19}$$

Our assumptions on the input data allow us to use the estimator in [16] for the covariance matrix (3.6):

$$\begin{aligned}
\tilde{\boldsymbol{C}}_i = &\frac{1}{\delta t}\mathbb{E}(\boldsymbol{y}_i(t+\delta t)\boldsymbol{y}_i(t+\delta t)|\boldsymbol{y}_i(t)) \\
&- \mathbb{E}(\boldsymbol{y}_i(t+\delta t)|\boldsymbol{y}_i(t))\mathbb{E}(\boldsymbol{y}_i(t+\delta t)|\boldsymbol{y}_i(t))
\end{aligned} \tag{3.20}$$

that according to [16] satisfies:

$$\boldsymbol{C}_i = \tilde{\boldsymbol{C}}_i + O(\delta t)$$

and by the Law Of Large Numbers we can estimate $\tilde{\boldsymbol{C}}_i$ using the empirical covariance of $\Delta \boldsymbol{y}_i(t_j)$:

$$\tilde{\boldsymbol{C}}_i = \frac{1}{M-1} \sum_{j=1}^{M-1} [\Delta \boldsymbol{y}_i(t_j) - \hat{\Delta y_i})][\Delta \boldsymbol{y}_i(t_j) - \hat{\Delta y_i})]^T$$

and get (3.19).

**Diffusion Maps.** For the purpose of finding a global parametrization that embodies the relation between the system variables, we use a kernel-based manifold learning technique called diffusion maps [18, 14]. This technique enables us to find a meaningful representation of the system state variables. In the sequel, we will briefly review the method in the context of our work.

Suppose that we have the features of $N$ system states, i.e $(\hat{\boldsymbol{z}}_i, \hat{\boldsymbol{C}}_i)$ for $i = 1, \ldots, N$, computed from the measurements. We denote by $\boldsymbol{W}$ the $N \times N$ pairwise affinity matrix between the features, whose $(i, l)$th element is given by:

$$\boldsymbol{W}_{i,l} = \exp\left\{ -\frac{d(\boldsymbol{z}_i, \boldsymbol{z}_l)}{4\epsilon} \right\}$$

where the (square) distance is

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) = \frac{1}{2} (\boldsymbol{z}_i - \boldsymbol{z}_l)^T (\boldsymbol{C}_i^{-1} + \boldsymbol{C}_l^{-1})(\boldsymbol{z}_i - \boldsymbol{z}_l) \tag{3.21}$$

and $\epsilon > 0$ is the kernel scale, usually set as the median of the pairwise distances. We define a corresponding diffusion operator $\mathbf{K}$ by:

$$\boldsymbol{K}_{i,l} = \frac{\boldsymbol{W}_{i,l}}{w(i)}$$

where

$$w(i) = \sum_{l=1}^{N} \boldsymbol{W}_{i,l} \tag{3.22}$$

Based on the spectral decomposition of $\boldsymbol{K}$ we build a global representation of the system states $\Psi$. Let $\lambda_0, ..., \lambda_{N-1}$ and $\boldsymbol{\psi}_0, ..., \boldsymbol{\psi}_{N-1}$ be the eigenvalues and eigenvectors of $\boldsymbol{K}$, respectively, written in descending order, so that $\lambda_{N-1} \leq ... \leq \lambda_0 = 1$. Using the $P$ eigenvectors corresponding to the largest $P$ eigenvalues, we define the following (nonlinear) map for each state $i$ into a $P$-dimensional space:

$$i \mapsto \boldsymbol{\Psi}_i = (\boldsymbol{\psi}^1(i), \boldsymbol{\psi}^2(i), ..., \boldsymbol{\psi}^P(i)) \in \mathbb{R}^P$$

This embedding of the data can be viewed as a new representation, which locally satisfies the main goal described in (3.4). We conclude this chapter with the presentation of the proposed algorithm in Algorithm 2.

**Algorithm 2** The Proposed Algorithm

**Input**: $M$ measurements of $N$ different states, i.e., $\boldsymbol{y}_i(t_j)_{j=1}^{M} \in \mathbb{R}^s,\ i = 1, \ldots, N$.
**Output**: A low dimensional representation of each state $\boldsymbol{\Psi}_i \in \mathbb{R}^P$.

1. For each state $i$, compute the feature $\hat{\boldsymbol{z}}_i$ and the covariance matrix $\hat{\boldsymbol{C}}_i$ by:

$$\hat{\boldsymbol{z}}_i = \frac{1}{M}\sum_{j=1}^{M} \boldsymbol{y}_i(t_j)$$

$$\hat{\boldsymbol{C}}_i = \frac{1}{M-1}\sum_{j=2}^{M}[\Delta\boldsymbol{y}_i(t_j) - \hat{\Delta\boldsymbol{y}_i}][\Delta\boldsymbol{y}_i(t_j) - \hat{\Delta\boldsymbol{y}_i}]^T.$$

   where

$$\hat{\Delta\boldsymbol{y}_i} = \frac{1}{M-1}\sum_{j=1}^{M} \Delta\boldsymbol{y}_i(t_j)$$

2. Build the pairwise affinity matrix $\boldsymbol{W}$ between all states by:

$$\boldsymbol{W}_{i,l} = \exp - \left\{ \frac{d(\boldsymbol{z}_i, \boldsymbol{z}_l)}{\epsilon} \right\}$$

   where

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) = \frac{1}{2}(\boldsymbol{z}_i - \boldsymbol{z}_l)^T(\boldsymbol{C}_i^{-1} + \boldsymbol{C}_l^{-1})(\boldsymbol{z}_i - \boldsymbol{z}_l)$$

3. Compute the diffusion operator $\boldsymbol{K}$ by:

$$\boldsymbol{K}_{i,l} = \frac{\boldsymbol{W}_{i,l}}{w(i)},\ w(i) = \sum_{l=1}^{N} W_{i,l}$$

4. Calculate the spectral decomposition of $\boldsymbol{K}$ and obtain its eigenvalues $\{\lambda_l\}_{l=0}^{N-1}$ and right eigenvectors $\{\boldsymbol{\psi}_l\}_{l=0}^{N-1}$.

5. Build a nonlinear mapping (embedding) of the system state:

$$(\hat{\boldsymbol{z}}_i, \hat{\boldsymbol{C}}_i) \mapsto \boldsymbol{\Psi}_i = (\boldsymbol{\psi}^1(i), \boldsymbol{\psi}^2(i), ..., \boldsymbol{\psi}^P(i))$$

# 4 Method Illustration

## 4.1 Simulation

To illustrate the proposed algorithm, consider the following evolution of the state variable:

$$\theta_i(t_{j+1}) - \theta_i(t_j) = -(\bar{\theta}_i - \theta_i(t_j))\Delta t + \sqrt{\Delta t} w_{i,\theta}$$

where $w_{i,\theta} \sim N(0, 0.09)$, $\Delta t = 0.05$, $1 \leq j \leq 250$ and the baseline values are given by

$$\bar{\theta}_i = \begin{cases} -5 & \text{for } 1 \leq i \leq 10 \\ 10 & \text{for } 11 \leq i \leq 20 \\ 50 & \text{for } 21 \leq i \leq 30 \end{cases}$$

An example of the evolution in time of the state variables $\theta_i(t_j)$ is shown in Figure 4.1 colored with respect to their baseline value.

In addition, consider the following evolution of the noise variable:

$$\eta_i(t_{j+1}) - \eta_i(t_j) = -(\bar{\eta}_i - \eta_i(t_j))\Delta t + \frac{1}{\epsilon}\sqrt{\Delta t} w_{i,\eta}$$

where $w_{i,\eta} \sim N(0, 0.09)$, $\epsilon = 0.1$, and the baseline values of the noise are uniformly sampled from $\bar{\eta}_i \in \{0, \ldots, 100\}$.

Suppose that the hidden variables $(\theta_i(t_j), \eta_i(t_j))$ are observed through the following non linear function:

$$\begin{aligned} \boldsymbol{y}_i(t_j) &= f(\theta_i(t_j), \eta_i(t_j)) \\ &= (\theta(t_j)^2 + 3\eta_i(t_j)^2, \theta_i(t_j)^2 - \eta_i(t_j)^2) \end{aligned}$$

We note that this nonlinear observation function $f$ satisfies Assumption 6. In Figure 4.2, we plot the measurements $y_i(t_j) \in \mathbb{R}^2$ for $i = 1, \ldots, 30$.

For each state $i$, we compute the features $(\hat{\boldsymbol{z}}_i, \hat{\boldsymbol{C}}_i)$ based on the measurements $\boldsymbol{y}_i(t_j)$ and apply Algorithm 2 with $P = 1$. In Figure 4.3, we display a comparison between the output of Algorithm 2, namely, $\boldsymbol{\psi}^1(i)$ and the true (inaccessible) baseline value $\bar{\theta}_i$

We observe that the computed one dimensional representation exhibits high correspondence with the hidden intrinsic state $\bar{\theta}_i$. Indeed, in this particular case, the true intrinsic state can be approximated simply by scaling the new representation, $\bar{\theta}_i \approx \boldsymbol{\psi}_i^1/\alpha$ for $\alpha = 0.005$.

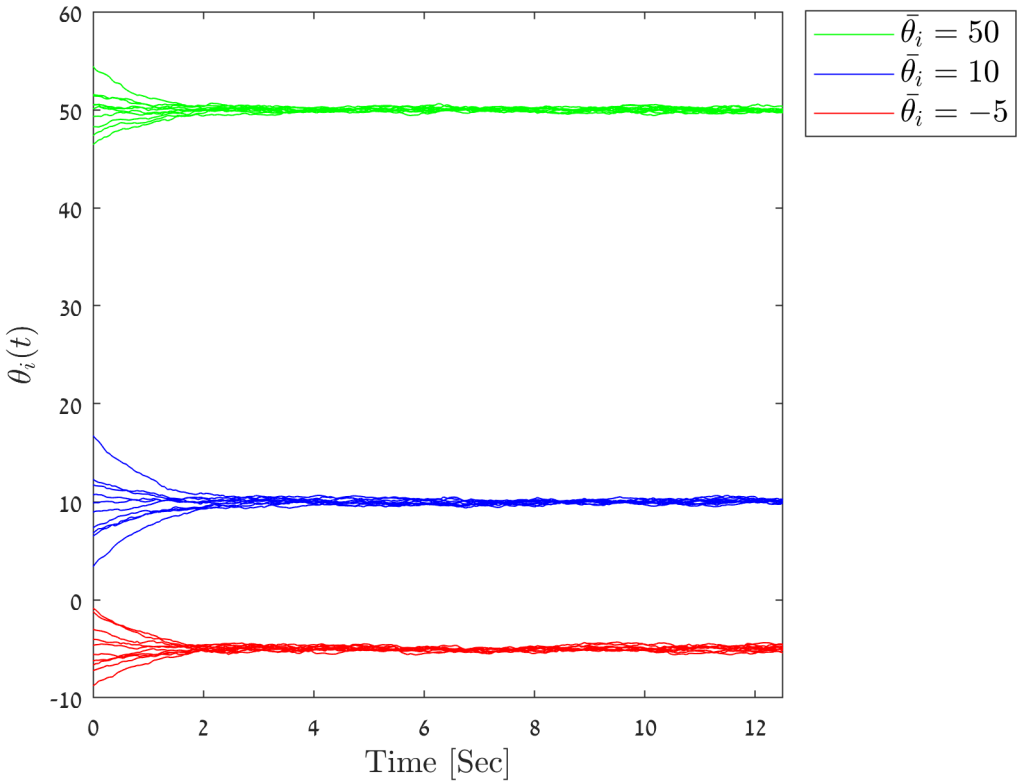



Figure 4.1: The evolution in time of the state variables $\theta_i(t_j)$ colored according to their baseline value $\bar{\theta}_i$

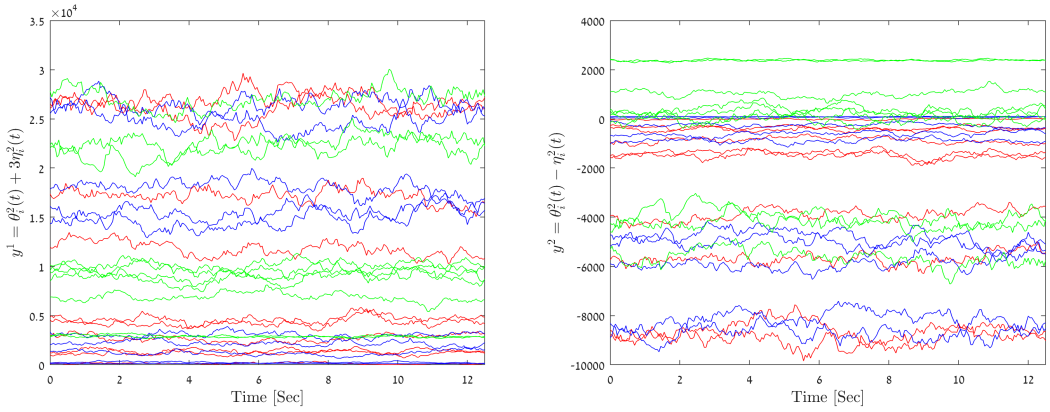
## 4.2 Toy Mechanical System

To further illustrate the proposed method, we apply Algorithm 2 to real measurements of a simple mechanical system. On the one hand, we show here the recovery of the main properties of the system from its observations in a data-driven manner without prior knowledge on the system. On the other hand, this particular mechanical system was chosed since it has a known definitive characterization, which can serve as a ground truth in our experiment to assess and validate the empirical results.

The mechanical system consists of two masses, $m_1$ and $m_2$ that are coupled with a spring with constant $k_2$. Each mass is connected to the ground with 2 additional springs with constant $k_1$.

Let $x_1$ and $x_2$ denote the position of the masses. An external force, denoted by $F1$, is applied to the mass $m_1$. A diagram of the mechanical system is depicted in Figure 4.4.

The experiment was conducted in repeated trials. In each trial, the two masses were set from a predefined grid consisting of 30 points, where $m_1 \in \{0, \ldots, 4\}$ and $m_2 \in \{0, \ldots, 5\}$. An external force (a square function) was used to invoke the system using a voice-coil actuator. With an optic-laser sensor, we measured the position of the mass $m_2$ in time. Each trial duration was 100 seconds and the sampling rate was 10 kHz. Let $g_i(t_j)$




(a) (b)

Figure 4.2: The measurements $y_i(t_j) \in \mathbb{R}^2$ for $i = 1, \ldots, 30$ colored according to the baseline value $\bar{\theta}_i$, where (a) depicts the first coordinate, and (b) depicts the second coordinate.

denote the time-series of the measured signal at the $i$th trial for $t_j = 1, \ldots, 1,000,000$.

In the context of the present work, we have observations of a mechanical system from 30 different states, where each state $i$ is specified by the values of the two masses $m_1$ and $m_2$.

Figure 4.5 shows an example of the system measurements from different states colored with respect to the sum of the masses.

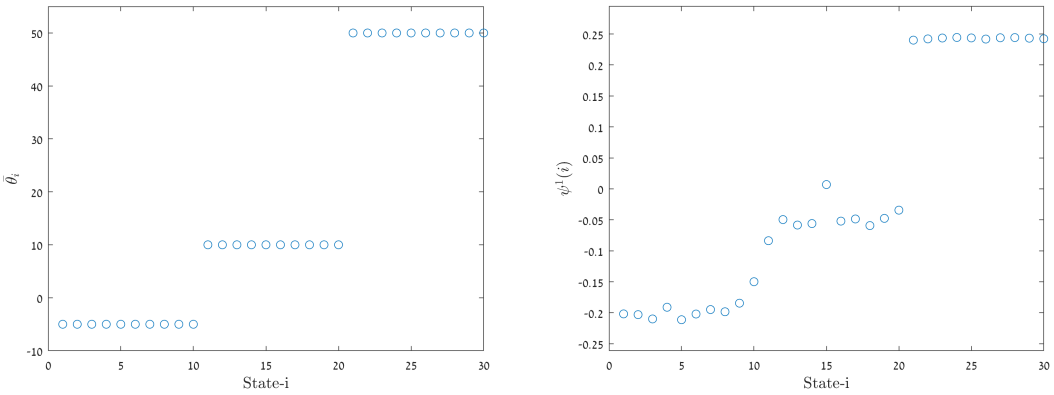
We follow common-practice in manifold learning and apply a pre-processing stage to the 1D time-series of the observations. Specifically here, we computed the spectrogram of each time series using an analysis window of length 1000 with overlap of 500. Let $\boldsymbol{y}_i(t_j) \in \mathbb{R}^s$ denote the resulting spectrogram at time $t_j = 1 \ldots, M$ in state $i = 1 \ldots, N$.

**System Analysis.** Using Newton's law, the ODE that describes the movement of each mass is given by:

$$\begin{aligned} m_1\ddot{x_1} &= F(t) - 2k_1x_1 - k_2(x_1 - x_2) - c_1\dot{x_1} \\ m_2\ddot{x_2} &= -2k_1x_2 - k_2(x_1 - x_2) - c_2\dot{x_2} \end{aligned}$$

We omit the dumping factor of each mass, namely, $c_1$ and $c_2$, and recast the ODEs in a matrix form:

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x_1} \\ \ddot{x_2} \end{bmatrix} + \begin{bmatrix} 2k_1 + k_2 & k_2 \\ k_2 & 2k_1 + k_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} F(t) \\ 0 \end{bmatrix}$$




(a) (b)

Figure 4.3: (a) The ground truth – the baseline value of the state variable as function of the state index. (b) The output of Algorithm 2 – the one dimensional representation as function of the state index.

The modes of the system can be found by solving an eigenvalue problem of the matrix $K^{-1}M$, where

$$M = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}$$
$$K = \begin{bmatrix} 2k_1 + k_2 & k_2 \\ k_2 & 2k_1 + k_2 \end{bmatrix}$$

The characteristic polynomial is:

$$\begin{aligned} \det(K^{-1}M - \lambda I) &= \det \begin{bmatrix} (2k_1 + k_2) \cdot m_1 - \lambda & -k_2 \cdot m_1 \\ -k_2 \cdot m_2 & (2k_1 + k_2) \cdot m_2 - \lambda \end{bmatrix} \\ &= [(2k_1 + k_2) \cdot m_1 - \lambda] \cdot [(2k_1 + k_2) \cdot m_2 - \lambda)] - k_2^2 \cdot m_1 \cdot m_2 \\ &= \lambda^2 - \lambda \cdot (2k_1 + k_2) \cdot (m_1 + m_2) - k_2^2 \cdot m_1 \cdot m_2 \end{aligned}$$

The system has two degrees of freedom, which are the roots of the characteristic polynomial, given by:

$$\begin{aligned} \lambda_{1,2} &= \frac{1}{2}\left[(2k_1 + k_2) \cdot (m_1 + m_2) \pm \sqrt{(2k_1 + k_2)^2 \cdot (m_1 + m_2)^2 + 4 \cdot k_2^2 \cdot m_1 \cdot m_2}\right] \\ &= \frac{1}{2}\left[(2k_1 + k_2) \cdot (m_1 + m_2) \pm (2k_1 + k_2) \cdot (m_1 + m_2)\sqrt{1 + \frac{4 \cdot k_2^2 \cdot m_1 \cdot m_2}{(2k_1 + k_2)^2 \cdot (m_1 + m_2)^2}}\right] \end{aligned}$$

Assuming that the springs remain constant during the experiment, we note that the two modes of the system are governed by the sum of the masses $m_1 + m_2$. This implies that the hidden state variable in each trial is $\bar{\theta}_i = m_1 + m_2$.

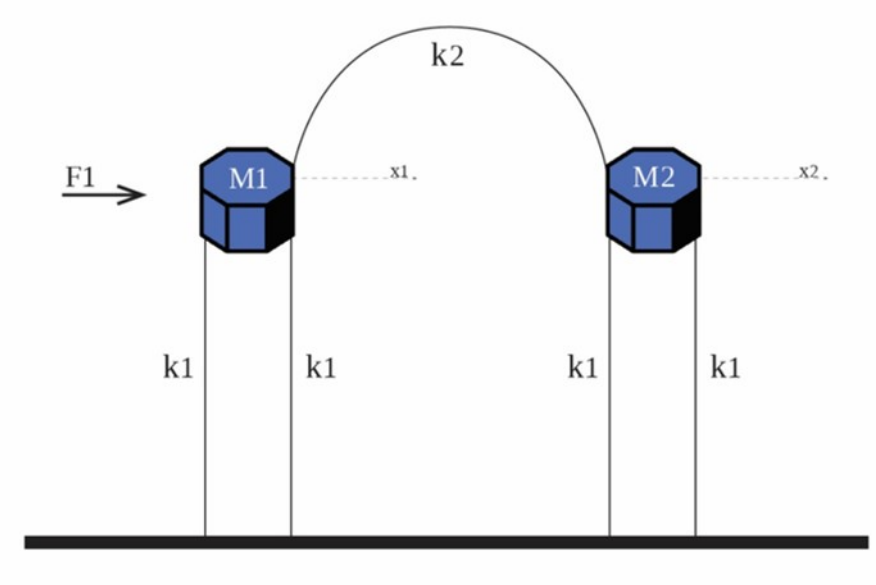



Figure 4.4: A diagram of the mechanical system.

**Results.** We apply Algorithm 2 to the input data $\{y_i(t_j)\}_{j=1}^{M}$. As a baseline, we apply a similar algorithm, where instead of using the modified Mahalanobis distance we use the Euclidean distance between the features $\boldsymbol{z}_i$ (in Step 2 of Algorithm 2, the affinity matrix $\boldsymbol{W}$ is computed using $\|\boldsymbol{z}_i - \boldsymbol{z}_l\|_2^2$ instead of $d(\boldsymbol{z}_i, \boldsymbol{z}_l)$ ). Figure 4.6 displays the 2D representation of the measurements resulting from the applications of the two algorithms. Figure 4.6a shows the results of Algorithm 2 and Figure 4.6b shows the results of the baseline algorithm. Each point in the figures represents a state (trial). The points are colored by the corresponding value of $m_1 + m_2$.

We observe that the 2D representation obtained by Algorithm 2 is organized according to the sum of the masses in each state, a result that is consistent with the analysis of the system presented above. Moreover, we observe that the contribution of using the modified Mahalanobis distance rather than the Euclidean distance between the features $\boldsymbol{z}_i$ is critical to the recovery of the true hidden state of the system.

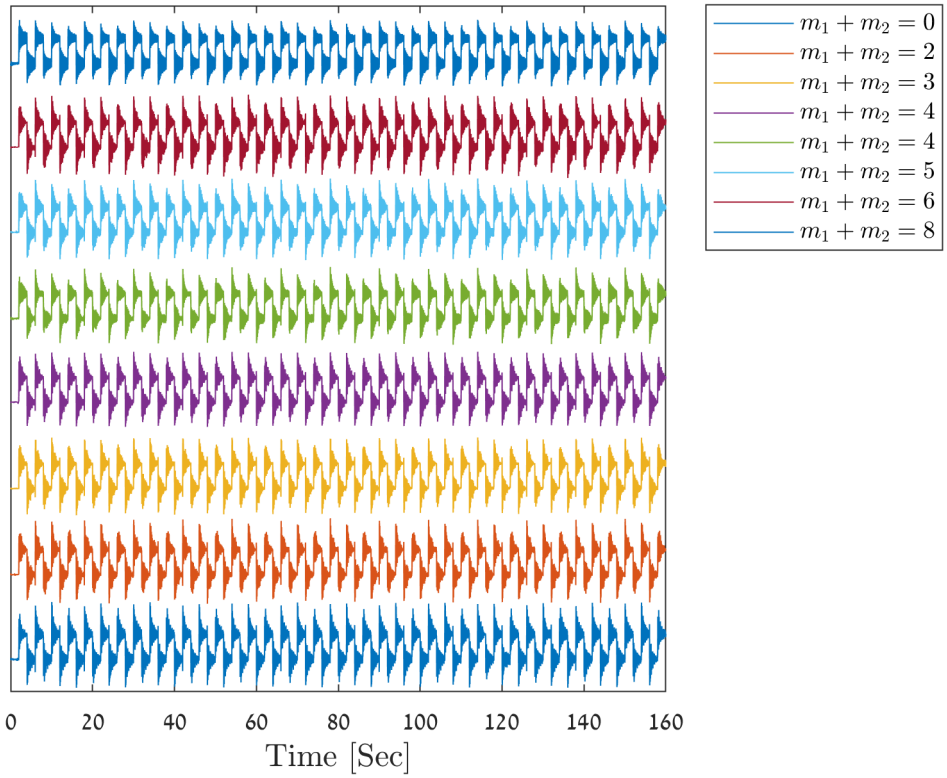
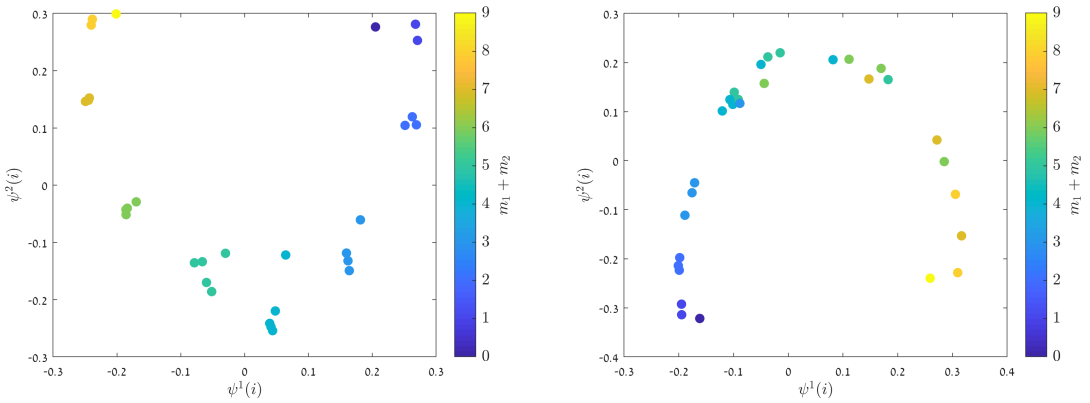



Figure 4.5: An example of the mechanical system measurements (the position of $m_2$) at some specific states as a function of time. The measurements are colored by the sum of the masses at each state.




(a) (b)

Figure 4.6: (a) The representation obtained by Algorithm 2 with $P = 2$. We present the scatter plot of the two most dominant eigenvectors, where the color depicts $m_1 + m_2$. (b) The representation obtained by the baseline algorithm.

# 5 Deep Brain Stimulation

Here, we utilize the method presented in Chapter 3 for unsupervised detection of target regions during a DBS surgery. We focus on two particular detection tasks: finding the subthalamic nucleus (STN) region and a sub-territory within the the STN region, called the dorsolateral oscillatory region (DLOR).

The measured signals are time series of neuronal activity at different depths along a pre-planned trajectory recorded by a micro-electrode. The time series at different depths are of varying lengths, depending on the recording time at each depth during the surgery. In the context of the problem setting described in Chapter 3, we refer to each specific depth as a system state, denoted by $i$ ,where $i = 1, .., N$. Accordingly, let $g_i(\tau_j), j = 1 \ldots, T_i$ be the time series of the signal recorded at depth $i$, where $\tau_j$ is the discrete time index and $T_i$ is the length of the signal recorded at depth $i$.

Similarly to the previous chapter, we apply a pre-processing stage to the 1D time series $g_i(\tau_j)$, where here we compute the Scattering Transform [26, 27], which is based on a cascade of wavelet transforms and modulus operators. Let $y_i(t_j)$ denote the resulting Scattering Transform at time $j = 1 \ldots, M_i$ in state $i = 1 \ldots, N$, where $M_i$ is the number of scattering transform time frames.

The signal acquired at each depth (state of the system) is classified by a human expert into one of four classes: Before STN, STN-DLOR, STN-Ventro Medial Non-oscillatory Region (VMNR), and After STN.

An illustrative example of the data is depicted in Figure 5.1, where we plot the time series measurements from 6 different depths (states), colored according to the expert labels.

We observe that signals within the STN region (colored in red and green) have higher variability compared to signals outside the STN region (colored in blue and cyan). Indeed, this variability was used as a feature in previous work, e.g. in [28] and [29, 30]. We also observe that there is no evident difference between the two classes of signals within the STN region, indicating that the DLOR detection is a challenging task.

## 5.1 Subthalamic Nucleus (STN) Detection

The proposed algorithm for the detection of the STN region appears in Algorithm 3, where we denote by $EDT(i)$ the $i$th coordinate of the Estimate Target from Distance (EDT) vector designating the specific depth along the pre-planned trajectory.

Note that our empirical examination suggests that the STN location is determined by the most dominant component, that is the eigenvector $\psi^1(i)$ corresponding to the largest eigenvalue. Therefore, the detection is based only on $\psi^1(i)$ resulting from the application

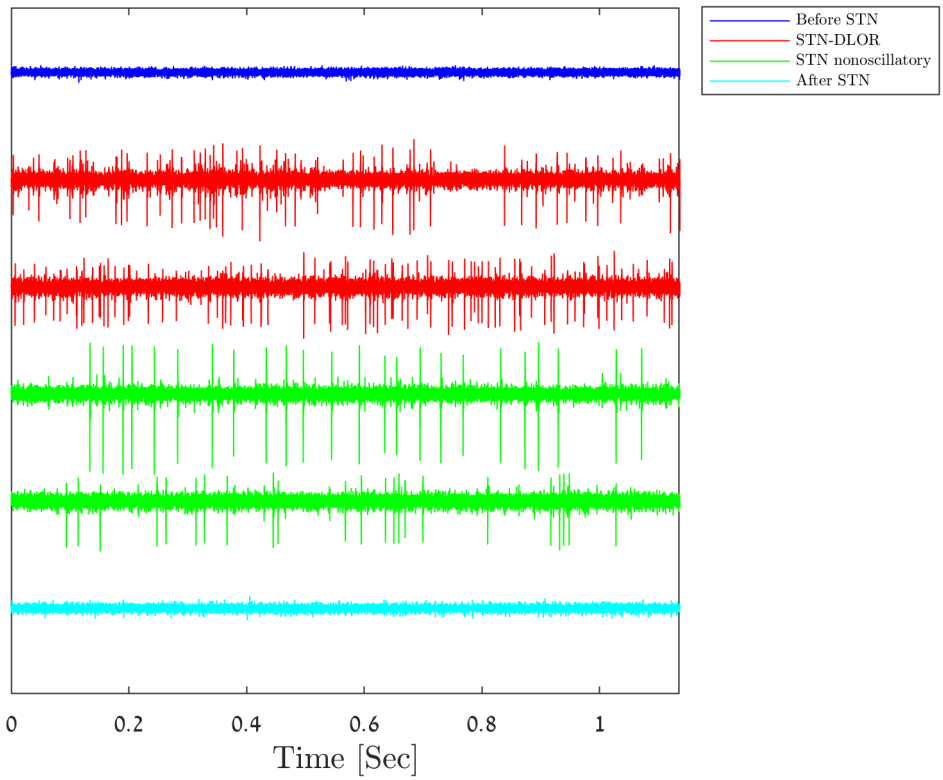
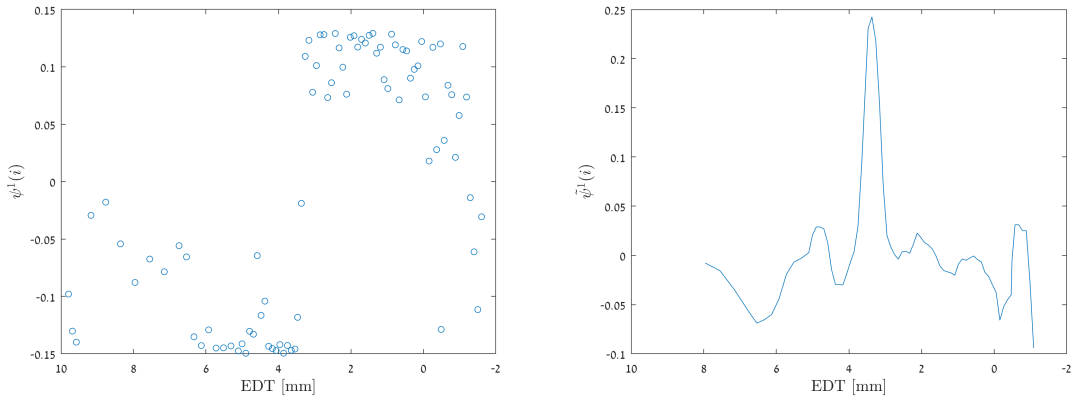



Figure 5.1: Time series measurements of the neuronal activity along the pre-planned trajectory colored according to the expert region labels: white matter before STN (blue), DLOR - DorsoLateral Oscillatory Region (DLOR) (red), Ventro Medial Non-oscillatory Region (VMNR)(green), and white matter after STN (light blue).

of Algorithm 2 to the pre-processed data. The detection itself is implemented in steps 3-5. The main idea is to detect the first sharp transition of values in $\psi^1(i)$, indicating the entrance to the STN region. In order to alleviate the effect of small perturbations, we smooth $\psi^1(i)$ by a moving average with a window of size 3 samples. Then, we detect the transition by computing the difference between the medians at two consecutive running windows of size 5 samples and obtain $\tilde{\psi}^1(i)$. The depth at which the maximal difference is obtained is set as the entrance point to the STN, denoted by $i_{en}$. Once the entrance point is determined, the exit point is set as the first point at which $\psi^1(i)$ is smaller than $\psi^1(i_{en})$.

The result of the application of Algorithm 3 to the example presented in Figure 5.1 is shown in Figure 5.2. In Figure 5.2a we plot $\psi^1(i)$ as a function of the depth $EDT(i)$. In Figure 5.2b, we plot $\tilde{\psi}^1(i)$ as a function of the depth $EDT(i)$.

It is important to note that the eigenvectors are always determined up to a sign. Therefore, in order to eliminate this inherent sign ambiguity, we replace $\psi^1(i)$ by $\text{sign}(\delta) \cdot \psi^1(i)$, where $\delta = |\max(\tilde{\psi}^1) - \tilde{\psi}^1(1)| - |\min(\tilde{\psi}^1) - \tilde{\psi}^1(1)|$.




(a) (b)

Figure 5.2: (a) 1D embedding obtained by the proposed method. The Y-axis displays the inferred state variable, which is the most dominant eigenvector of the kernel generated by the proposed method as a function of Estimated Target from Distance (EDT). (b) The difference between medians of consecutive running windows of size 5 samples along the pre-defined trajectory. The Y-axis displays the most dominant eigenvector after processing (moving average of size 3 samples and calculating the median difference) as a function of the EDT.

## 5.2 Dorsolateral Oscillatory Region (DLOR) Detection

The proposed algorithm for the DLOR detection appears in Algorithm 4. Since the transitions between the sub-territories of the STN region are subtle and not as distinct as the transition in and out of the STN, we perform two adjustments with respect to Algorithm 3. First, we assume that the information on subtle changes in the system's state variables is manifested deeper in the spectrum, namely in smaller eigenvalues. Therefore, we use more than one coordinate (eignevectors) to embed the measurements. Second, we use a small prior on the data – that the STN region is divided into continuous regions. Accordingly, we modify the diffusion operator as follows:

$$\boldsymbol{K}^t = \boldsymbol{K} + \boldsymbol{K}^s,$$

where $\boldsymbol{K}$ is the same operator used for the detection of the STN region, and $\boldsymbol{K}^s$ is a smoothing kernel that emphasizes relations between consecutive depths, given by:

$$W^s_{i,l} = \exp\left\{-\frac{||d_s(i) - d_s(j)||^2}{\epsilon}\right\}$$

$$K^s_{i,l} = \frac{W^s_{i,l}}{w^s(i)} \quad , \quad w^s(i) = \sum_{l=1}^{N} W^s_{i,l}$$

**Algorithm 3** STN Region Detection

**Input**: $g_i(\tau_j) \in \mathbb{R}^{T_i}, i = 1, .., N$ – Time series measurements of neuronal activity at different depths,
$EDT$ – vector indicating the Estimated Distance from Target of each depth.
**Output**: *STN entrance point* and *STN exit point*.

1. For each time series $g_i(\tau_j)$, compute its scattering transform:

$$\{y_i(t_j)\}_{j=1}^{M_i} = \Phi(g_i(\tau_j)) \in \mathbb{R}^k, i = 1, .., N$$

2. Compute $\psi^1(1)$ according to Algorithm 2

3. Compute $\tilde{\psi^1}$ by applying a moving average with a window of size 3 samples to $\psi^1$, and then, compute the difference the medians at two consecutive running windows of size 5 samples

4. $i_{en} = \underset{i}{\text{argmax}} \quad \tilde{\psi^1}(i)$

5. $i_{ex} = \underset{i}{\text{argmin}} \quad \{i : \psi^1(i) \leq \psi^1(i_{en}) \quad \text{and} \quad i > i_{en}\}$

6. $STN\ entrance\ point = EDT(i_{en})$, $STN\ exit\ point = EDT(i_{ex})$

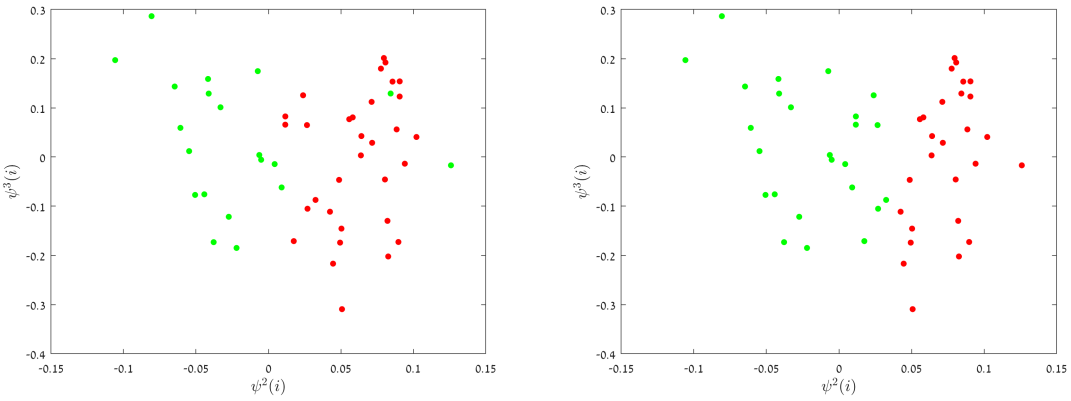
where $d_s(i)$ is the $i$th coordinate of the $EDT$ vector which states the specific depth along the trajectory. This way we encourage smoother transitions in the embedded space and increase the robustness of the diffusion operator $\boldsymbol{K}^t$ to noise.

According to the above adjustments, we apply eigenvalue decomposition to $\mathbf{K}^t$ and represent each depth in the STN region with *two* eigenvectors corresponding to the two largest eigenvalues, excluding $\psi_0$ and $\psi_1$, since $\psi_0$ is trivial and $\psi_1$ contains information on the STN boundaries rather on the sub-territories within the STN.

This representation enables us to find a separation in the embedded space. Consequently, in order to complete the task of an unsupervised detection of the DLOR region, we apply K-means [31] to the embedded representation of each depth with an additional coordinate – the state's depth. We apply k-means with $k = 2$ initialized with the entrance depth and exist depth of the STN region.

An example of the 2D embedding of measurements from depths within the STN region is shown in Figure 5.3. In Figure 5.3a, the points are colored according to the DLOR labels obtained by a human expert, where red points belong to depths within the DLOR and green points belong to depths labeled as STN-VMNR. The corresponding K-means labels are displayed in Figure 5.3b.

Finally, based on Algorithm 3 and 4, we cluster the data according to the 4 labels. A visual comparison between the labels obtained by our unsupervised method and the supervised HMM algorithm with respect to labels given by an expert to data from a




(a) (b)

Figure 5.3: (a) The 2D embedding of signals at different depths along the pre-planned trajectory colored according to the expert labels, where red points reside in the DLOR and green points reside in the VMNR according to the expert labels. In the 2D embedding, the X-axis indicates the second most dominant eigenvector and the Y-axis indicates the third most dominant eigenvector of the kernel computed by the proposed method. (b) The 2D embedding colored according to k-means applied in Algorithm 4. The red points reside in the DLOR and green points reside in the VMNR according to algorithm 4.

specific example is presented in Figure 5.4, where we denote our method by USVA (Unsupervised State Variables Approximation).

## 5.3 Quantitative Detection Results

We apply our proposed method (USVA), particularly Algorithm 3 and Algorithm 4 to 25 different trajectories recorded from 16 patients, and we compare the results to the results obtained by the algorithm proposed in [28], which is considered the gold-standard.

Each trajectory consists of three transition points of interest: the STN entrance, the DLOR exit, and the STN exist. In order to evaluate the detection, we define an objective measure as the distance between the transition point marked by the human expert and the detected transition point. For the purpose of normalization, we divide the distance by the size of the respective region. Consequently, we have a total of six performance measures: STN entrance and exit errors (divided by the size of the STN region), DLOR entrance and exit errors (divided by the size of the DLOR region), and the overall entrance and exit errors. Note that the STN entrance error and the DLOR entrance error differ only by the normalization factor, since the STN entrance point coincides with the DLOR entrance point.

The median and interquartile range (IQR) of all the performance measures are re-

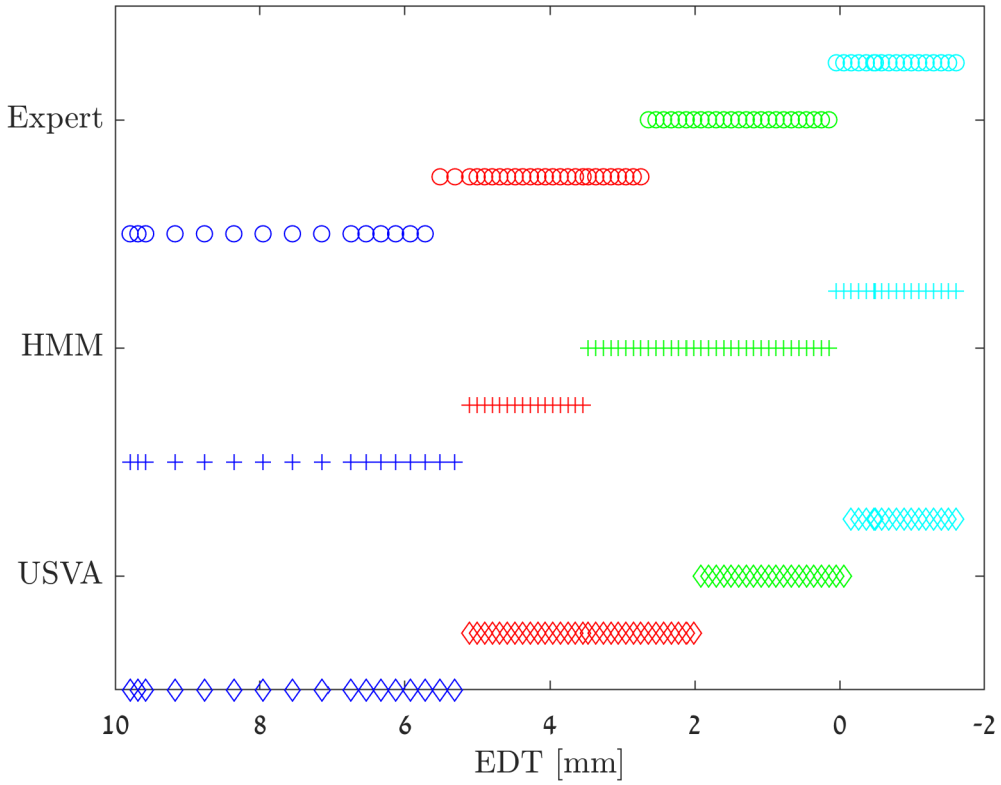



Figure 5.4: An illustration of the detection results attained by our method, the HMM algorithm, and the expert labels along the pre-planned trajectory. The figure displays the labels given by each method as a function of the EDT. Each point is colored with respected to the label, as in Figure 5.1, where points marked by 'o' are the expert label, points marked by '+' are the HMM label, and points marked by '□' are our labels.

ported in Figure 5.5. To complement the experimental study, we also report their mean and standard deviation, which are as follows. The mean STN entrance error in percentage obtained by our algorithm is $4.77 \pm 6.97$ compared to $4.41 \pm 8.62$ achieved by the HMM. The mean STN exist error in percentage obtained by our algorithm is $3.95 \pm 5.53$ compared to $2.89 \pm 4.56$ achieved by the HMM. The mean STN overall error in percentage obtained by our algorithm is $8.72 \pm 11.05$ compared to an average of $7.31 \pm 9.8$ achieved by the HMM. The mean DLOR exist error in percentage obtained by our algorithm is $19.21 \pm 8.48$ compared to $39.16 \pm 34.87$ achieved by the HMM. The mean DLOR overall error in percentage obtained by our algorithm is $25.81 \pm 11.6$ compared to $43.86 \pm 37.03$ achieved by the HMM. We remark that in our performance evaluation, failures to detect the DLOR exist point is considered as a 100% error.

We observe that our method attains results comparable to the gold-standard in the detection of the STN region and outperforms the gold-standard in the detection of the DLOR.

**Algorithm 4** DLOR Detection

**Input**: the affinity matrix $\boldsymbol{K} \in \mathbb{R}^{N\times N}$, the STN entrance point, the STN exit point $d_s = EDT$ – the vector indicating the Estimated Distance from Target of each depth.
**Output**: *DLOR exit point.*

1. Compute a smoothing kernel:

$$W^s_{i,l} = \exp\left\{-\frac{||d_s(i) - d_s(j)||^2}{\epsilon}\right\}, \quad K^s_{i,l} = \frac{W^s_{i,l}}{w^s(i)}, \quad w^s(i) = \sum_{l=1}^{N} W^s_{i,l}$$

2. Compute the kernel:

$$\boldsymbol{K}^t = \boldsymbol{K} + \boldsymbol{K}^s$$

3. Apply the eigenvalue decompositio to $\boldsymbol{K}^t$ and obtain its eigenvalues and eigenvectors.

4. Represent each depth in the STN region according to the eigenvectors associated with the third and fourth highest eigenvalues, i.e $R(i) = [\psi^2(i), \psi^3(i), d_s(i)]$

5. Cluster all depths representation $R$ into 2 clusters according to K-means initialized with the entrance and exit points.

6. $i_d = \underset{i}{\text{argmin}}\{i : R(i) \in STN \quad non - oscillatory\}$

7. *DLOR exit point* $= EDT(i_d)$

## 5.4 Globus Pallidus (GP) Detection

We further illustrate the generality of our method by demonstrating a proof-of-concept of a different target detection task during a DBS surgery. Specifically, we utilize our proposed in order to find a region called Globus Pallidus (GP). We remark that this is a region of interest in DBS surgeries aimed at treating dystonia, whereas the STN is commonly of interest in DBS surgeries aimed at treating Parkinson's disease.

The setting of the signals as well as the pre-processing are similar to the STN detection case, and therefore, the measured signals and notation are the same. Namely, $g_i(\tau_j)$ denotes the measured time series of neuronal activity at depth (state) $i$, and $y_i(t_j)$ denotes the resulting Scattering Transform at time $j = 1..., M_i$.

Here, the signals at each depth are classified by a human expert into one of four classes: before GP – Striatum (str), GPe, GPi, and after GP (exit). An illustrative example of the data is depicted in Figure 5.6, where we plot the time series measurements from 6 different depths, colored according to the expert labels. We can observe that there is no distinct characteristic of the different regions, and therefore, discriminating between them in an unsupervised manner is a challenging task.

Similarly to the procedure described in Section 5.1, we compute the affinity matrix $\boldsymbol{K}$ as in Algorithm 2 and a smoothing kernel $\boldsymbol{K}^s$ as in Algorithm 4. Then, we construct the joint kernel $\boldsymbol{K}^t = \boldsymbol{K} + \boldsymbol{K}^s$, and based on the spectral components of $\boldsymbol{K}^t$, we represent each depth. The 2D embedding of a single trajectory defined by the two most dominant eigenvectors of $\boldsymbol{K}^t$, $\psi_1$ and $\psi_2$, is depicted in Figure 5.7. In Figure 5.7a the points are colored by the depth along the pre-defined trajectory and in Figure 5.7b by the expert labels.

We observe that the clusters in Figure 5.7a coincide with the expert labels. As in the STN detection task, the current gold-standard is a supervised HMM algorithm [32]. An illustration of the detection along a specific trajectory is presented in Figure 5.8.

We observe that the detection of the transitions between the Striatum and the GPe and between the GPe and the GPi achieved by our unsupervised method are more consistent with the labels of the expert compared to the supervised HMM algorithm.

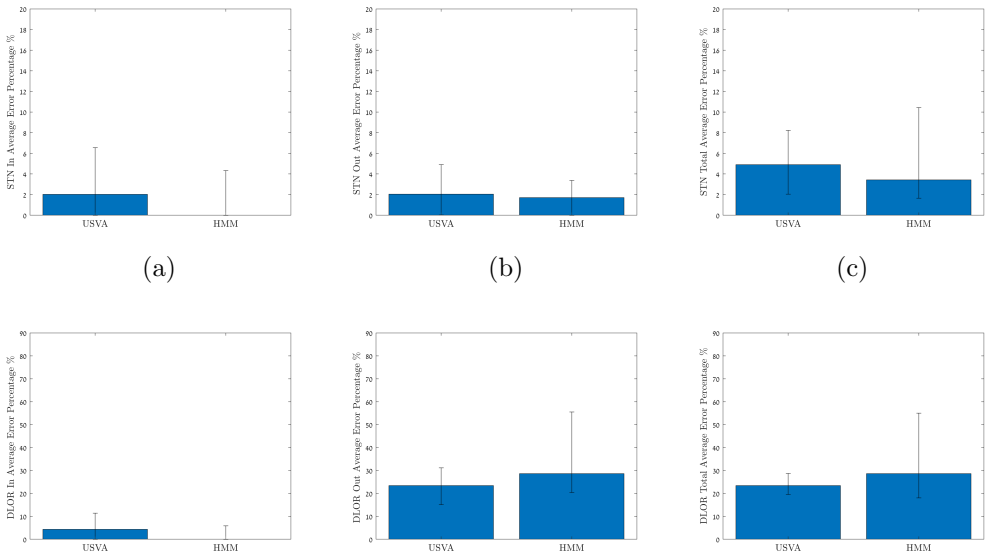



Figure 5.5: Performance comparison between our method (USVA) and the HMM-based algorithm proposed in [28], which is considered the gold-standard. The comparison is made with respect to the labels given by an expert to 25 different trajectories acquired from 16 patients. The bars indicate the median error percentage and the confidence intervals indicate its interquartile range (IQR). At the top row, we present the performance of the two competing algorithms (Algorithm 3 and the algorithm proposed in [28]) in detecting the STN region, and at the bottom row, we present the performance of the two algorithms (Algorithm 4 and the algorithm proposed in [28]) in detecting the DLOR. (a) The error in detecting the entrance to the STN region. (b) The error in detecting the exit from the STN region (c) The overall error in detecting the STN region. (d) The error in detecting the entrance to the DLOR. (e) The error in detecting the exit from the DLOR. (f) The overall error in detecting the DLOR.

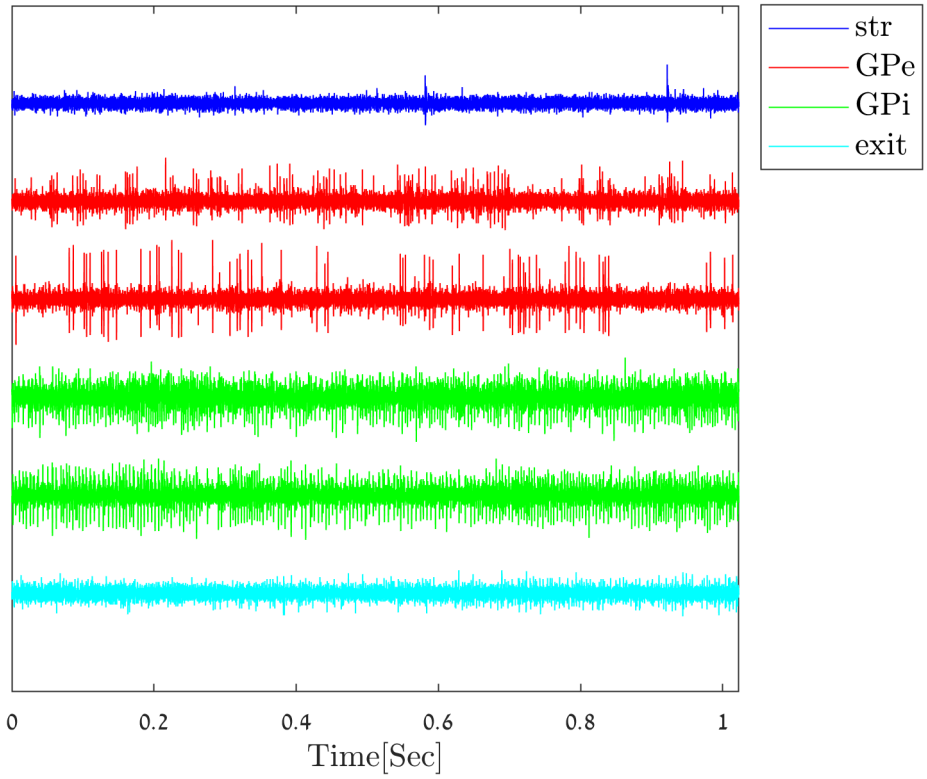
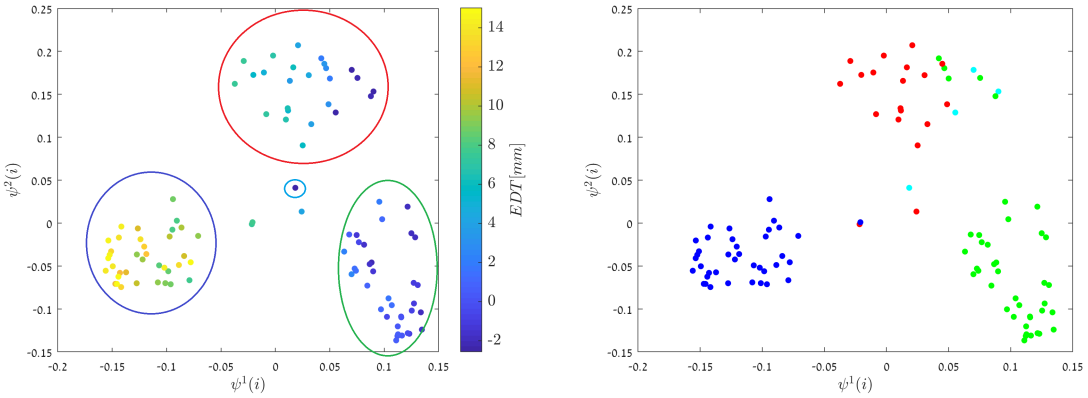



Figure 5.6: Time series measurements of the neuronal activity along the pre-planned trajectory colored according to the expert region labels: Before GP (str) in blue, GPe in red, GPi in green, and After GP (exit) in light blue.




(a) (b)

Figure 5.7: The 2D embedding defined by the two most dominant eigenvectors. (a) The color indicates the trajectory depth, and the circles mark the obtained clusters. (b) The color indicates the expert labels as in Figure 5.6.

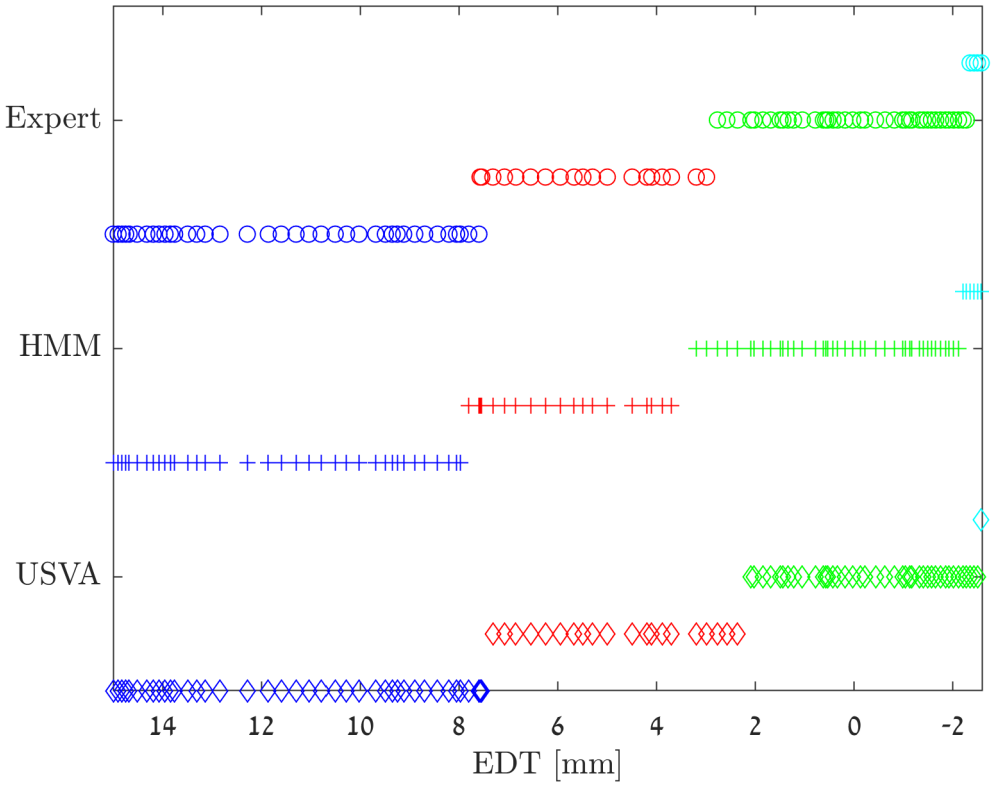



Figure 5.8: An illustration of the detection results attained by our method, the HMM algorithm, and the expert labels along the pre-planned trajectory. The figure displays the labels given by each method as a function of the EDT. Each point is colored with respected to the label, as in Figure 5.6, where points marked by 'o' are the expert label, points marked by '+' are the HMM label, and points marked by '□' are our labels.

# 6 Conclusion and Future work

In this research, we introduced a new unsupervised method for target and anomaly detection based on manifold learning.

In Chapter 3, we formulated the problem setup. Our focus was on a problem comprising multiple sets of measurements of a dynamical system acquired at different system states, where each set contains repeated measurements from the same state. We assume that the measurements are governed by two types of latent variables. The first type of variables is *state variables* that control the intrinsic behaviour of the system. Our assumption is that these variables have a prototypical value per state with only small perturbations around it. The second type of variables is *noise variables*, and our assumption is that they are characterized by high variation with respect to the *state variables*.

Considering the above problem, in Chapter 3, we propose to use a modified version of the classical Mahalanobis distance as a distance function between specially-designed features computed at each state from the available measurements. We showed that the proposed distance facilitates the separate of the two types of latent variables, namely state variables and noise variables, by giving rise to an approximation of the Euclidean distance between the *state variables* alone, thereby implicitly ignoring the *noise variables*. By incorporating this distance in the diffusion maps framework, we devised a data driven method for extracting the state variables, which enables us to represent the intrinsic system state. In turn, this intrinsic representation is used as a proxy for the desired target detection.

In Chapter 4, we demonstrated the proposed method. First we presented a simulation, where we showed that our method is able to recover the latent state variables even in adverse noisy environments (with negative SNRs). Then, we validated our method by applying it to real measurements of a mechanical system, where we showed that our method gives rise to a representation that coincides with the (known and available) theoretical analysis of this mechanical system.

In Chapter 5, we presented two specifications of our method for particular target detection tasks involving DBS. Based on our method, we developed unsupervised algorithms for the detection of specific target regions in the brain, called the STN and a sub-territory with in the STN called the DLOR. The detection of these two regions is of great interest during DBS surgeries for alleviating the symptoms of Parkinson's disease. We showed that our algorithm outperforms the current gold standard in detecting the DLOR and obtains comparable results to the gold standard in the detection of the STN, even though the current gold standard is a supervised method and our approach is unsupervised. Finally, we presented a proof-of-concept of the application of our method to the detection of another brain area, called the GP, which is of interest during a different

DBS surgery.

We believe that this work can serve as a solid stepping-stone for several research directions. One future research direction is to further the study of GP detection, devising a suitable unsupervised algorithm with a comprehensive analysis and experimental study, including comparisons to the current gold standard [32]. Another research direction stems from the fact that we need to compute specially-tailored features based on the measurements without any prior knowledge, and then to estimate their covariance. Finding better estimators for the covariance matrix based on the data, for example using shrinkage [33, 34], may significantly improve the method results and enable us to extend the scope to other applications. Perhaps the most significant future research direction involves the generalization of the proposed method to multiple sets of measurements from different modalities. In the context of manifold learning, multimodal data fusion has attracted much attention recently, e.g. [35, 36]. We intend to incorporate the proposed variant of the Mahalanobis distance for the purpose of devising unsupervised target and anomaly detection methods for multi-channel, multi-modal data.

# Bibliography


[1] Varun Chandola, Arindam Banerjee, and Vipin Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 15, 2009.

[2] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 260, pp. 2319–2323, 2000.

[3] Sam T. Roweis and Lawrence K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 260, pp. 2323–2326, 2000.

[4] David L. Donoho and Carrie Grimes, "Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 5591–5596, 2003.

[5] Mikhail Belkin and Parth Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural. Comput.*, vol. 15, no. 6, pp. 1373–1396, June 2003.

[6] Ronen Talmon, Stéphane 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.

[7] Or Yair, Ronen Talmon, Ronald R Coifman, and Ioannis G Kevrekidis, "Reconstruction of normal forms by learning informed observation geometries from data," *Proceedings of the National Academy of Sciences*, vol. 114, no. 38, pp. E7865–E7874, 2017.

[8] Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[9] Amit Singer, Yoel Shkolnisky, and Boaz Nadler, "Diffusion interpretation of nonlocal neighborhood filters for signal denoising," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 118–139, 2009.

[10] Rubén Ibanez, Emmanuelle Abisset-Chavanne, Jose Vicente Aguado, David Gonzalez, Elias Cueto, and Francisco Chinesta, "A manifold learning approach to data-driven computational elasticity and inelasticity," *Archives of Computational Methods in Engineering*, vol. 25, no. 1, pp. 47–57, 2018.

[11] Tal Shnitzer, Mirela Ben-Chen, Leonidas Guibas, Ronen Talmon, and Hau-Tieng Wu, "Recovering hidden components in multimodal data with composite diffusion operators," *SIAM Journal on Mathematics of Data Science*, vol. 1, no. 3, pp. 588–616, 2019.

[12] Hau-tieng Wu, Ronen Talmon, and Yu-Lun Lo, "Assess sleep stage by modern signal processing techniques," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1159–1168, 2014.

[13] Ronen Talmon and Ronald R Coifman, "Empirical intrinsic geometry for nonlinear modeling and time series filtering," *Proceedings of the National Academy of Sciences*, vol. 110, no. 31, pp. 12535–12540, 2013.

[14] Ronald R Coifman and Stéphane Lafon, "Diffusion maps," *Applied and computational harmonic analysis*, vol. 21, no. 1, pp. 5–30, 2006.

[15] Amit Singer and Ronald R. Coifman, "Non-linear independent component analysis with diffusion maps," vol. 25, no. 2, pp. 226 – 239, 2008.

[16] Carmeline J Dsilva, Ronen Talmon, C William Gear, Ronald R Coifman, and Ioannis G Kevrekidis, "Data-driven reduction for a class of multiscale fast-slow stochastic dynamical systems," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 3, pp. 1327–1351, 2016.

[17] Bernt Oksendal, *Stochastic differential equations: an introduction with applications*, Springer Science & Business Media, 2013.

[18] Ronald. R. Coifman and Stéphan Lafon, "Diffusion maps," vol. 21, no. 1, pp. 5–30, 2006.

[19] Ronen Talmon and Ronald R. Coifman, "Empirical intrinsic geometry for nonlinear modeling and time series filtering," *Proc. Nat. Acad. Sci.*, vol. 110, no. 31, pp. 12535–12540, 2013.

[20] Ronen Talmon and Ronald R Coifman, "Intrinsic modeling of stochastic dynamical systems using empirical geometry," *Applied and Computational Harmonic Analysis*, vol. 39, no. 1, pp. 138–160, 2015.

[21] Amit Singer, Radek Erban, Ioannis G Kevrekidis, and Ronald R Coifman, "Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps," *Proceedings of the National Academy of Sciences*, vol. 106, no. 38, pp. 16090–16095, 2009.

[22] Dimitrios Giannakis, "Data-driven spectral decomposition and forecasting of ergodic dynamical systems," *Applied and Computational Harmonic Analysis*, vol. 47, no. 2, pp. 338–396, 2019.

[23] Ronen Talmon, Dan Kushnir, Ronald R Coifman, Israel Cohen, and Sharon Gannot, "Parametrization of linear systems using diffusion kernels," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1159–1173, 2011.

[24] Carmeline J Dsilva, Ronen Talmon, C William Gear, Ronald R Coifman, and Ioannis G Kevrekidis, "Data-driven reduction for a class of multiscale fast-slow stochastic dynamical systems," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 3, pp. 1327–1351, 2016.

[25] Gregory F Lawler, "Stochastic calculus: An introduction with applications," *American Mathematical Society*, 2010.

[26] Stéphane Mallat, "Group invariant scattering," *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012.

[27] Joan Bruna, Stéphane Mallat, Emmanuel Bacry, Jean-François Muzy, et al., "Intermittent process analysis with scattering moments," *The Annals of Statistics*, vol. 43, no. 1, pp. 323–351, 2015.

[28] Dan Valsky, Odeya Marmor-Levin, Marc Deffains, Renana Eitan, Kim T Blackwell, Hagai Bergman, and Zvi Israel, "S top! border ahead: A utomatic detection of subthalamic exit during deep brain stimulation surgery," *Movement Disorders*, vol. 32, no. 1, pp. 70–79, 2017.

[29] Adam Zaidel, Alexander Spivak, Lavi Shpigelman, Hagai Bergman, and Zvi Israel, "Delimiting subterritories of the human subthalamic nucleus by means of microelectrode recordings and a hidden markov model," *Movement disorders*, vol. 24, no. 12, pp. 1785–1793, 2009.

[30] Stephen Wong, GH Baltuch, JL Jaggi, and SF Danish, "Functional localization and visualization of the subthalamic nucleus from microelectrode recordings acquired during dbs surgery with unsupervised machine learning," *Journal of neural engineering*, vol. 6, no. 2, pp. 026006, 2009.

[31] John A Hartigan and Manchek A Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[32] Dan Valsky, Kim T Blackwell, Idit Tamir, Renana Eitan, Hagai Bergman, and Zvi Israel, "Real-time machine learning classification of pallidal borders during deep brain stimulation surgery," *Journal of Neural Engineering*, vol. 17, no. 1, pp. 016021, 2020.

[33] Matan Gavish, Ronen Talmon, Pei-Chun Su, and Hau-Tieng Wu, "Optimal recovery of mahalanobis distance in high dimension," *arXiv preprint arXiv:1904.09204*, 2019.

[34] David L Donoho, Matan Gavish, and Iain M Johnstone, "Optimal shrinkage of eigenvalues in the spiked covariance model," *Annals of statistics*, vol. 46, no. 4, pp. 1742, 2018.

[35] Roy R Lederman and Ronen Talmon, "Learning the geometry of common latent variables using alternating-diffusion," *Applied and Computational Harmonic Analysis*, vol. 44, no. 3, pp. 509–536, 2018.

[36] Moshe Salhov, Ofir Lindenbaum, Yariv Aizenbud, Avi Silberschatz, Yoel Shkolnisky, and Amir Averbuch, "Multi-view kernel consensus for data analysis," *Applied and Computational Harmonic Analysis*, vol. 49, no. 1, pp. 208–228, 2020.

# זיהוי מטרות ואנומליות בעזרת למידת יריעה עם ישום בניתוח גירוי מוחי עמוק

עידו כהן

# זיהוי מטרות ואנומליות בעזרת למידת יריעה עם ישום בניתוח גירוי מוחי עמוק

חיבור על מחקר

לשם מילוי חלקי של הדרישות
לקבלת התואר מגיסטר למדעים בהנדסת חשמל

עידו כהן

הוגש לסנט הטכניון - מכון טכנולוגי לישראל
אפריל 2020 חיפה אייר תש"פ

# תודות

מחקר זה נעשה תחת הנחייתו של פרופסור רונן טלמון בפקולטה להנדסת חשמל ע"ש אנדרו וארנה ויטרבי. אנו מודים לטכניון על תרומתו הכספית הנדיבה שאפשרה את מחקר זה.

# עיקרי העבודה

בעיית זיהוי מטרות ואנומליות מוגדרת כמציאת תבניות במידע או תבניות שחורגות מהתנהגות מצופה. בעיה זו נחשבת למאתגרת במיוחד מכיוון שלתבניות הללו עשויה להיות שונות גדולה, ואנומליות, כמעט בהגדרה, הן מאוד נדירות. בנוסף, כאשר אנו מודדים מערכת דינמית, המדידות מושפעות ממספר רב של משתנים, כשלרוב רק חלק קטן ממשתנים אלו אכן מגדירים את מצב המערכת, ואילו שאר המשתנים מוסיפים רעש והפרעות למדידות. מרבית האלגוריתמים הקיימים לזיהוי מטרות ואנומליות משתמשים בידע מוקדם על הבעיה, כגון הנחת מודל מסוים או שימוש במידע מתויג. עובדה זו יכולה לגרום להטיה וכתוצאה מכך ביצועי הזיהוי תלויים במידה רבה באיכות הידע המוקדם.

בכדי להימנע מתלות זו, פיתחנו שיטה לא מונחת המבוססת על למידת יריעה אשר מצליחה להפריד בין המשתנים המהותיים שמגדירים את מצב המערכת והמשתנים הזניחים שמייצגים הפרעות ורעש. על סמך שיטה זו, הצענו אלגוריתם לזיהוי מטרות ואנומליות שמבוסס אך ורק על המדידות. האלגוריתם שהצענו מבוסס על וקטור מאפיינים שמגלמים בתוכם את המידע על המערכת, וניתן לחשבם מהמדידות בלבד ללא ידע מוקדם. וקטור מאפיינים זה תוכנן במיוחד עבור פונקצית מרחק המבוססת על מרחק מהלנוביס אשר מדגישה את המשתנים המהותיים. את וקטור המאפיינים ופונקצית המרחק שפיתחנו אנו משלבים בשיטת למידת יריעה שנקראת מיפוי דיפוזיה, אשר בונה ייצוג למשתנים המהותיים. אנו מנתחים את השיטה המוצעת בעזרת חשבון סטוכסטי )איטו( ומראים כי פונקציית המרחק המוצעת בין וקטורי המאפיינים אכן חושפת בקירוב טוב את המשתנים המהותיים של המידע, ובכך מסייעת לנו לבנות אלגוריתם מדויק לזיהוי מטרות ואנומליות.

אנחנו מדגימים את השיטה שלנו על שני ישומים. ראשית, אנחנו מדגימים את השימוש בשיטה על מדידות של מערכת מכנית פשוטה. בחרנו במערכת מכנית של שתי מטוטלות מצומדות מכיוון שקיים עבורה מודל ידוע אשר משמש אותנו לאימות הייצוג שמניבה השיטה המוצעת. ואכן, אנו מראים שמאפייני המערכת שחולצו על ידי השיטה המוצעת באופן לא מונחה וללא ידע מקדים, אכן מתלכדים עם המאפיינים התיאורטיים של המערכת.

הישום השני עוסק בבעיית זיהוי מטרה במהלך ניתוח גירוי מוחי עמוק. גירוי מוחי עמוק הוא טיפול שמשגר אותות חשמליים בעזרת קוצב למקום מסוים במוח שאחראי על מנגנון התנועה בגוף. לאחר שהקוצב מותקן במקום המתאים במוח, הוא עוזר להפחית תסמינים שנגרמים ממחלות נוירולוגיות כגון פרקינסון או דיסטוניה. דוגמאות לתסמינים כאלה הן רעידות לא רצוניות של חלקי גוף (בעיקר גפיים או לסת), קשיחות או חוסר גמישות (קישיון) של המפרקים, אטיות, ומיעוט או מחסור בתנועה.
במהלך הניתוח אחת המשימות המרכזיות היא למצוא את האזור המתאים במוח להשתלת הקוצב. הגילוי נעשה בעזרת אלקטרודה המודדת את הפעילות העצבית במוח באזורים שונים. לרוב מדידות אלו רועשות מאוד וחילוץ המידע הרלוונטי מהם אינו טריוויאלי. אנו מתמקדים בניתוח שנועד להקל על התסמינים של מחלת הפרקינסון. במהלך ניתוח זה נדרש לזהות אזור במוח הנקרא "הגרעין התת־תלמי" ותת אזור בתוכו המכונה האזור האוסילטורי. זיהוי מדויק של אזורים אלה הוא הכרחי לתוצאות קליניות רצויות. על בסיס השיטה שפיתחנו במחקר זה, אני מציעים אלגוריתם לא מונחה לזיהוי האזור התת תלמי ותת האזור בתוכו במהלך ניתוח גירוי מוח עמוק על סמך מדידות של הפעילות העצבית. אנו מראים כי השיטה שלנו משיגה תוצאות דומות לשיטה המקובלת כיום לזיהוי הגרעין התת־תלמי ואף מראה ביצועים משופרים בזיהוי האזור האוסילטורי. חשוב לציין כי השיטה המקובלת כיום היא שיטה מונחת המנצלת תיוג של מומחה, ואילו השיטה המוצעת על ידינו אינה מוטה לתיוג של מומחה מסוים.
בנוסף לישומים לעיל, אנו מציגים הוכחת התכנות לזיהוי של גרעין בסיס במוח הנקרא "הגרעין החיוור" (הגלובוס פלידוס) אשר משמש לטיפול בדיסטוניה. הוכחת התכנות זו מדגימה את ההיקף הרחב של הישומים האפשריים של השיטה המוצעת, ובמיוחד, ישומים שהם חסרים תיוגים אמינים.