Unsupervised Detection of Sub-Territories of the Subthalamic Nucleus During DBS Surgery With Manifold Learning

Ido Cohen, Dan Valsky, and Ronen Talmon, *Senior Member, IEEE*

***Abstract*—During Deep Brain Stimulation (DBS) surgery for treating Parkinson's disease, detecting the Subthalamic Nucleus (STN) and its sub-territory called the Dorsolateral Oscillatory Region (DLOR) is crucial for adequate clinical outcomes. Currently, the detection is based on human experts, often guided by supervised machine learning detection algorithms. This procedure depends on the knowledge and experience of particular experts and on the amount and quality of the labeled data used for training the machine learning algorithms. In this paper, to circumvent such dependence and the inevitable bias introduced by the training data, we present a data-driven unsupervised algorithm for detecting the STN and the DLOR during DBS surgery based on an agnostic modeling approach. Given measurements, we extract new features and compute a variant of the Mahalanobis distance between these features. We show theoretically that this distance enhances the differences between measurements with different intrinsic characteristics. Incorporating the new features and distance into a manifold learning method, called Diffusion Maps, gives rise to a representation that is consistent with the underlying factors that govern the measurements. Since this representation does not rely on rigid modeling assumptions and is obtained solely from the measurements, it facilitates a broad range of detection tasks; here, we propose a specification for STN and DLOR detection during DBS surgery. We present detection results on 25 sets of measurements recorded from 16 patients during surgery. Compared to a supervised algorithm, our unsupervised method demonstrates similar results in detecting the STN and superior results in detecting the DLOR.**

***Index Terms*—Deep brain stimulation, diffusion maps, Mahalanobis distance, manifold learning, Parkinson's disease, subthalamic nucleus.**

Manuscript received 8 March 2022; revised 15 July 2022 and 12 September 2022; accepted 2 October 2022. Date of publication 17 October 2022; date of current version 21 March 2023. This work was supported in part by the Technion Hiroshi Fujiwara Cyber Security Research Center and in part by the Pazy Foundation under Grant 78-2018. The work of Ronen Talmon supported by Schmidt Career Advancement Chair in AI. *(Corresponding authors: Ido Cohen; Dan Valsky.)*

Ido Cohen is with the Andrew and Erna Viterby Faculty of Electrical and Computer Engineering, Technion Israel Institute of Technology, Haifa 3200003, Israel (e-mail: sidoc@campus.technion.ac.il).

Ronen Talmon is with the Andrew and Erna Viterby Faculty of Electrical and Computer Engineering, Technion Israel Institute of Technology, Israel.

Dan Valsky is with the The Edmond and Lily Safra Center for Brain Research, The Hebrew University of Jerusalem, Jerusalem, Israel, and also with the Center of Functionally Integrative Neuroscience, Aarhus University, 8000 Aarhus, Denmark (e-mail: dan@cfin.au.dk).

Digital Object Identifier 10.1109/TBME.2022.3215092

I. Introduction

DEEP Brain Stimulation (DBS) is a treatment involving an implanted stimulating device that sends electrical signals to brain areas that are responsible for body movements. Once the device is implanted in an appropriate position, DBS can help to reduce the symptoms of tremor, slowness, stiffness, and walking problems caused by several neuronal diseases, such as Parkinson's disease, dystonia, or essential tremor. More specifically, we focus on the DBS of the Subthalamic Nucleus (STN), which is a known and effective treatment for Parkinson's disease [1], [2]. During the surgical procedure to implant the DBS lead, one important task is to detect the exact location of the STN borders and a sub-territory within it, called Dorsolateral Oscillatory Region (DLOR). In [3], [4], [5], [6], it was shown that accurate detection of these regions contributes substantially to the clinical benefit of STN-DBS.

The common procedure to detect the STN borders and the DLOR consists of two steps. First, before the surgery, a coarse approximation of the STN location is obtained based on Magnetic Resonance Imaging (MRI) and computed tomography (CT) images. This coarse approximation facilitates the determination of a pre-planned trajectory to the target region. Second, during the surgery, the exact detection of the target region is based on Micro Electrode Recordings (MERs) of neuronal activity along the pre-planned trajectory. The MERs are typically intricate and one needs to extract the relevant information for the detection of the target regions. In general, most target detection algorithms use prior knowledge on the data, such as predefined models, hand-crafted features, and human expert labels, in order to identify specific patterns that are indicative of the target region. For example, existing methods rely on various features extracted from the MERs, including the total power of the signal [7], the oscillatory activity in the beta (13–30 Hz) frequency band [8], [9], [10], and the high-frequency (> 500 Hz) neuronal "noise" [11], [12]. Then, based on such features, as well as on human expert labels, supervised classifiers are applied [13], [14], [15].

Current detection methods suffer from the following inherent shortcomings. First, the accuracy of supervised classifiers depends on the amount of labeled data, which is typically small in STN detection applications. Second, the hand-crafted features that are used in the detection do not necessarily carry sufficient information on the STN's exact location. Third, the labels are tagged by human experts, and therefore, naturally, are biased toward their specific experience and knowledge.

In this work, to alleviate such shortcomings we develop a data-driven unsupervised method and apply it to the STN and DLOR detection tasks performed during DBS surgery. Our method is based on an agnostic modeling approach that assumes the measurements typically include many sources of variability, where only a few of them are informative and facilitate the detection of specific target regions of interest. We assume that the measurements are the output of some unknown measurement function of hidden variables, which represent the sources of variability. We further assume that the hidden variables can be divided into two classes. The first class consists of intrinsic variables, which are characterized by slow dynamics. The second class consists of interference and noise variables, which are manifested by high measurement variances. Our method consists of three parts. In the first part, we present useful features that can be computed from the measurements. In the second part, we design a distance function between the features that capture the difference between the (slow) intrinsic variables and is invariant to the (fast) noise variables. Finally, in the third part, by making use of Diffusion Maps [16], we construct a global representation of the measurements and show that it is consistent with the intrinsic variables. Our premise, which we support by empirical evidence, is that the discovery of the intrinsic variables is useful for modeling, in general, and for detecting regions of interest, in particular.

The remainder of this paper is organized as follows. In Section II, we present the proposed method in a general context, and show both theoretically and in experiments that the proposed method is able to reveal the intrinsic representation of the measurements. In Section III, we propose a specification of the method for the problem of STN and DLOR border detection, yielding two purely unsupervised algorithms. We present the detection results obtained by our algorithm and compare its performance to existing supervised Hidden Markov Model (HMM) based algorithms [13]. Finally, in Section IV, we discuss the results and outline a few potential directions for future research.

II. Method - Unsupervised State Variables Approximation

Our method is based on a setting consisting of measurements of a stochastic dynamical system, obtained through some unknown observation function. The propagation model of the dynamical system is unknown, and the main assumption is that the system is driven by a set of intrinsic variables. In this section, we present an agnostic algorithm that builds a new representation of these intrinsic variables from the system measurements. The new embedding of the system's intrinsic variables facilitates accurate target detection of different system regimes. We start with a description of the general setup, and then we present our method and demonstrate it using simulations and real measurements from a simple mechanical system. Finally, we show a theoretical justification for our derivations. In Section III, we show a utilization of the algorithm for the STN and DLOR detection tasks.

A. Problem Formulation

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

$$d\boldsymbol{\theta}_i(t) = -\nabla U_{\bar{\theta}_i}(\boldsymbol{\theta}_i(t))dt + I_{d_1 \times d_1} d\boldsymbol{w}_{i,\theta}(t), \tag{1}$$

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

The second source of variability is considered to be noise, represented by 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:

$$d\boldsymbol{\eta}_i(t) = -\nabla U_{\bar{\eta}_i}(\boldsymbol{\eta_i}(t))dt + \frac{1}{\epsilon} I_{d_2 \times d_2} d\boldsymbol{w}_{i,\eta}(t), \tag{2}$$

where the drift is the gradient of a quadratic potential function given by $U_{\bar{\eta}_i}(\boldsymbol{\eta}) = \frac{1}{2}(\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}_i)^\top(\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}_i)$, $\bar{\boldsymbol{\eta}}_i$ is an unknown baseline constant, $I_{d_2 \times d_2}$ is a $d_2 \times d_2$ identity matrix, $\boldsymbol{w}_{i,\eta}(t)$ is a vector of d_2 independent Brownian motions, and $0 < \epsilon \ll 1$. Note that the variance of the diffusion term of the noise variables in (2) is larger than the variance of the diffusion term of the state variables in (1) by a factor of $1/\epsilon^2$. We assume that $\boldsymbol{w}_{i,\theta}(t)$ and $\boldsymbol{w}_{i,\eta}(t)$ are independent. We remark that (1) and (2) establish a prototypical propagation model of multi-scale stochastic dynamical systems [17], [18], [19].

Suppose that 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 smooth bi-Lipschitz, possibly nonlinear, function and $d = d_1 + d_2$.

For notational convenience, we denote

$$\boldsymbol{x}_i(t) = \begin{bmatrix} \boldsymbol{\theta}_i(t) \\ \boldsymbol{\eta}_i(t) \end{bmatrix},$$

and accordingly, we recast (1) and (2) as the following Itô process in d dimensions:

$$d\boldsymbol{x}_i(t) = -\nabla U(\boldsymbol{x}_i(t))dt + \Lambda d\boldsymbol{w}_i(t), \tag{3}$$

where $U(\boldsymbol{x})$ is a quadratic potential function centered at $\bar{\boldsymbol{x}}_i$, and

$$\bar{\boldsymbol{x}}_i = \begin{bmatrix} \bar{\boldsymbol{\theta}}_i \\ \bar{\boldsymbol{\eta}}_i \end{bmatrix}, \Lambda = \begin{bmatrix} I_{d_1 \times d_1} & 0 \\ 0 & \frac{1}{\epsilon} I_{d_2 \times d_2} \end{bmatrix}, \ \boldsymbol{w}_i(t) = \begin{bmatrix} \boldsymbol{w}_{i,\theta}(t) \\ \boldsymbol{w}_{i,\eta}(t) \end{bmatrix}.$$

We assume that we have access to M measurements from each state sampled in time on a discrete uniform grid. Accordingly, let $\boldsymbol{y}_i(t_j) \in \mathbb{R}^s$ denote the jth time measurement of the system at state i, where $t_j = \delta t \cdot j, j = \{0, 1, 2, \ldots, M-1\}$, and δt is the sampling time interval. Our goal is to decouple the two sources of variability, given the measurements $\boldsymbol{y}_i(t_j)$ without prior knowledge of the system variables, and to build an embedding of the system measurements that is consistent with the state variables. Since the state variables can be viewed as a proxy of the true state of the system, the ability to extract them may facilitate the identification of particular desired system regimes and target states. In the sequel, we will show how such an embedding sets the stage for accurate, unbiased STN and DLOR detection.

The specification of this 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 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

$$\boldsymbol{x}_i(t_j) = \begin{bmatrix} \boldsymbol{\theta}_i(t_j) \\ \boldsymbol{\eta}_i(t_j) \end{bmatrix}.$$

We aim to build an embedding of the measurements $\boldsymbol{y}_i(t_j)$, which represent the system state, and thereby, to identify in a purely unsupervised manner the STN region. In order to accomplish this goal, we devise a pairwise distance between system states that satisfies:

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) \approx \alpha ||\bar{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}}_l||^2, \tag{4}$$

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

B. Proposed Algorithm

We propose an unsupervised algorithm that is able to reveal the relations between the system's intrinsic variables without any prior knowledge. In this sub-section we focus on the utilization of our method relying on the analysis presented in Section II-E. Broadly, the proposed algorithm consists of two main stages. First, we represent each measurement by a set of features that can be computed solely from measurements and devise a distance function that achieves (4). Second, we apply a manifold learning method, Diffusion Maps, that constructs a global representation of the hidden state variables based on the proposed features and distance function.

1) Features and Distance Function: First, for each set of measurements $\{\boldsymbol{y}_i(t_j)\}_{j=1}^M$, we define features that can be computed solely from the measurements:

$$\hat{\boldsymbol{z}}_i = \frac{1}{M} \sum_{j=1}^{M} \boldsymbol{y}_i(t_j) \tag{5}$$

$$\hat{C}_i = \frac{1}{M-1} \sum_{j=2}^{M} [\boldsymbol{\mu}_i(t_j) - \hat{\boldsymbol{\mu}}_i)][\boldsymbol{\mu}_i(t_j) - \hat{\boldsymbol{\mu}}_i)]^\top \tag{6}$$

where we denote the increments between consecutive measurements by $\boldsymbol{\mu}_i(t_j) = \boldsymbol{y}_i(t_j) - \boldsymbol{y}_i(t_{j-1})$ and $\hat{\boldsymbol{\mu}}_i = \frac{1}{M-1} \sum_{j=2}^{M} \boldsymbol{\mu}_i(t_j)$, so that $\hat{\boldsymbol{z}}_i$ is the empirical mean of the measurements and $\hat{C}_i$ is the empirical covariance of the measurement increments.

Second, we define a metric between these features, enabling us to reveal the state variables. Particularly, we propose to use the following modified version of the (squared) Mahalanobis distance [20], [21]:

$$d(\hat{\boldsymbol{z}}_i, \hat{\boldsymbol{z}}_l) = \frac{1}{2} (\hat{\boldsymbol{z}}_i - \hat{\boldsymbol{z}}_l)^\top (\hat{C}_i^{-1} + \hat{C}_l^{-1})(\hat{\boldsymbol{z}}_i - \hat{\boldsymbol{z}}_l). \tag{7}$$

One of the main assumptions underlying our work is that the latent state variables $\boldsymbol{\theta}_i(t)$ are characterized by small perturbations around some informative baseline value, whereas the noise variables $\boldsymbol{\eta}_i(t)$ are characterized by high variance. In previous work, e.g., in [20] and [19], it was shown that this variant of the Mahalanobis distance implicitly attenuates hidden components with high variance without prior knowledge, motivating its utilization in our setting as well. In Section II-E, we show theoretically that the proposed distance attenuates the influence of the noise variables and gives rise to a distance between the hidden state variables. In Section II-D and Section III, we present empirical results that further support the usage of this distance.

2) Diffusion Maps: Manifold learning is a class of nonlinear geometry-oriented dimensionality reduction methods [22], [23], [24], [25]. 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 [16], [26]. Typically in manifold learning, a high dimensional data set that is assumed to lie on a low dimensional manifold is given. This class of methods attempts 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 [27], [28], earth structure classification [21], image reconstruction [29], signal denoising [30], numerical simulation enhancement [31], fetal electrocardiogram analysis [32], sleep stage identification [33], and time series filtering [34], to name but a few. 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., $(\hat{\boldsymbol{z}}_i, \hat{C}_i)$ for $i = 1, \ldots, N$, computed from the measurements. We denote by W the $N \times N$ pairwise affinity matrix between

Algorithm 1: 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 $\hat{C}_i$ according to (5) and (6).
2) Build the pairwise affinity matrix W between all states according to (8) and (7).
3) Compute the diffusion operator K according to (9)
4) Calculate the spectral decomposition of 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:

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

the features, whose (i, l)th element is given by:

$$W_{i,l} = \exp\left\{-\frac{d(\hat{\boldsymbol{z}}_i, \hat{\boldsymbol{z}}_l)}{\epsilon}\right\}, \tag{8}$$

where the (squared) distance is defined in (7), and $\epsilon > 0$ is the kernel scale, usually set as the median of the pairwise distances. We define a corresponding diffusion matrix K by:

$$K_{i,l} = \frac{W_{i,l}}{\boldsymbol{w}(i)}, \qquad \boldsymbol{w}(i) = \sum_{l=1}^{N} W_{i,l} \tag{9}$$

We remark that several different normalizations of the affinity matrix W were proposed in [26] and in related literature. In our work, we tested different normalizations and constructed K as presented since it yielded the best empirical results.

Based on the spectral decomposition of K, we build a global representation of the system states as follows. Let $\lambda^0, \ldots, \lambda^{N-1}$ and $\boldsymbol{\psi}^0, \ldots, \boldsymbol{\psi}^{N-1}$ be the eigenvalues and eigenvectors of K, respectively, written in descending order, so that $\lambda_{N-1} \leq \ldots \leq \lambda_0 = 1$. Using the P eigenvectors corresponding to the largest P eigenvalues, we define the following (nonlinear) map for each state i to a P-dimensional space:

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

Since this embedding of the data is based on an affinity that is locally invariant to the noise variables, i.e., the corresponding distance satisfies (4) (see Section II-E), we view it as a new representation of the hidden state variables. We conclude this section with a presentation of the proposed method in Algorithm 1.

C. Simulation Results

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

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

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

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

A realization of all the trajectories of the state variable $\theta_i(t_j)$ is shown in Fig. 1(a). In addition, we consider the following evolution of the noise variable:

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

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 $f : \mathbb{R}^2 \to \mathbb{R}^2$:

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

We note that this nonlinear observation function f satisfies Assumption 1 that is presented in Section II-E. In Fig. 1(b)–(c), we plot the two coordinates of the measurements $\boldsymbol{y}_i(t_j)$ in $\mathbb{R}^2$ for $i = \{1, \ldots, 30\}$. For each state i, we computed the features $\boldsymbol{z}_i$ and $\boldsymbol{C}_i$ based on the measurements $\boldsymbol{y}_i(t_j)$ and applied Algorithm 1. In Fig. 1(d)–(f), we display a comparison between the output of Algorithm 1 for $P = 1$, namely, $\boldsymbol{\psi}^1(i)$, and the true (inaccessible) baseline value $\bar{\theta}_i$.

We can observe that the computed one dimensional embedding has high correspondence with the hidden intrinsic baseline state $\bar{\theta}_i$, which can be approximated by a linear function: $\boldsymbol{\phi}_i^1 \approx \alpha\bar{\theta}_i, \quad \alpha = 0.005$, thereby achieving our main goal.

We remark that, in this specific example, our empirical results suggest that the intrinsic state of the system can be captured solely by the first leading eigenvector. However, in general, the information on the intrinsic state is often manifested in the first few leading eigenvectors. Therefore, subsequent high-dimensional clustering, such as k-means, is typically applied to the first leading eigenvectors. Importantly, such a generic clustering stage does not consider the temporal order of the samples, which is exploited in our algorithms presented in Section III.

D. Experimental Results on a Mechanical System

To further demonstrate the proposed method, we apply Algorithm 1 to real measurements of a 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 of the system. On the other hand, this particular mechanical system was chosen since it has a known definitive characterization, which can serve as ground truth in our experiment in order to assess and validate the empirical results.

The mechanical system we consider 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 two additional springs, each with constant k_1. Let x_1 and x_2 denote the positions of the two masses. An external force, denoted by $F1$, is applied to the

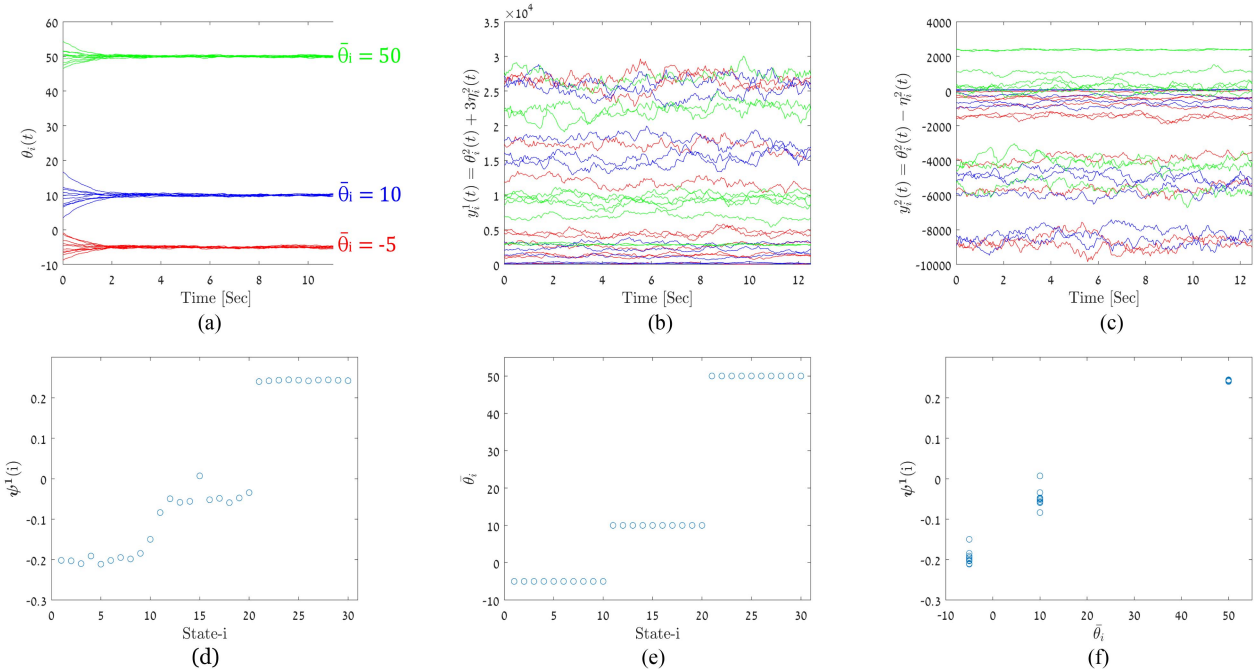

Fig. 1. Illustration of our method on simulations. (a) The system's hidden state variable $\theta_i(t)$ colored according to the baseline value $\bar{\theta}_i$. (b)–(c) The system's measurements $\boldsymbol{y}_i(t) = f(\theta_i(t), \eta_i(t)) \in \mathbb{R}^2$ colored according to their (hidden) baseline values $\bar{\theta}_i$. These measurements serve as the input data to our algorithm. (f) The one dimensional embedding $\psi^1(i)$ obtained by our suggested algorithm as a function of the state index. (e) The true baseline values $\bar{\theta}_i$ ("the ground truth") as a function of the state index. (f) A scatter plot that present the relation between the embedding obtained by our suggested algorithm and the "ground truth".

mass m_1. A diagram of the mechanical system is presented in Fig. 2(b).

The experiment comprised repeated trials. In each trial, the values of 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 wave 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 over time. The duration of each trial was 100 seconds and the sampling rate was 10 kHz. Let $g_i(t_j)$ denote the time-series of the measured signal at the ith trial for $t_j = 1, \ldots, 1,000,000$.

In analogy to our setting, 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.

Fig. 2(a) shows an example of the system measurements from different states, colored by the sum of the masses.

We follow common-practice in manifold learning and apply a pre-processing stage to the one dimensional time-series of observations. Specifically, we computed the spectrogram of each time series using an analysis window of length 1000 with an 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 $i = 1 \ldots, N$.

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

$$m_1\ddot{x}_1 = F(t) - 2k_1x_1 - k_2(x_1 - x_2) - c_1\dot{x}_1$$
$$m_2\ddot{x}_2 = -2k_1x_2 - k_2(x_1 - x_2) - c_2\dot{x}_2.$$

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

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix} + \begin{bmatrix} 2k_1 + k_2 & k_2 \\ k_2 & 2k_1 + k_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} F(t) \\ 0 \end{bmatrix}$$

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 corresponding characteristic polynomial is:

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

implying that the system has two degrees of freedom (the roots of the characteristic polynomial), which are given by:

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

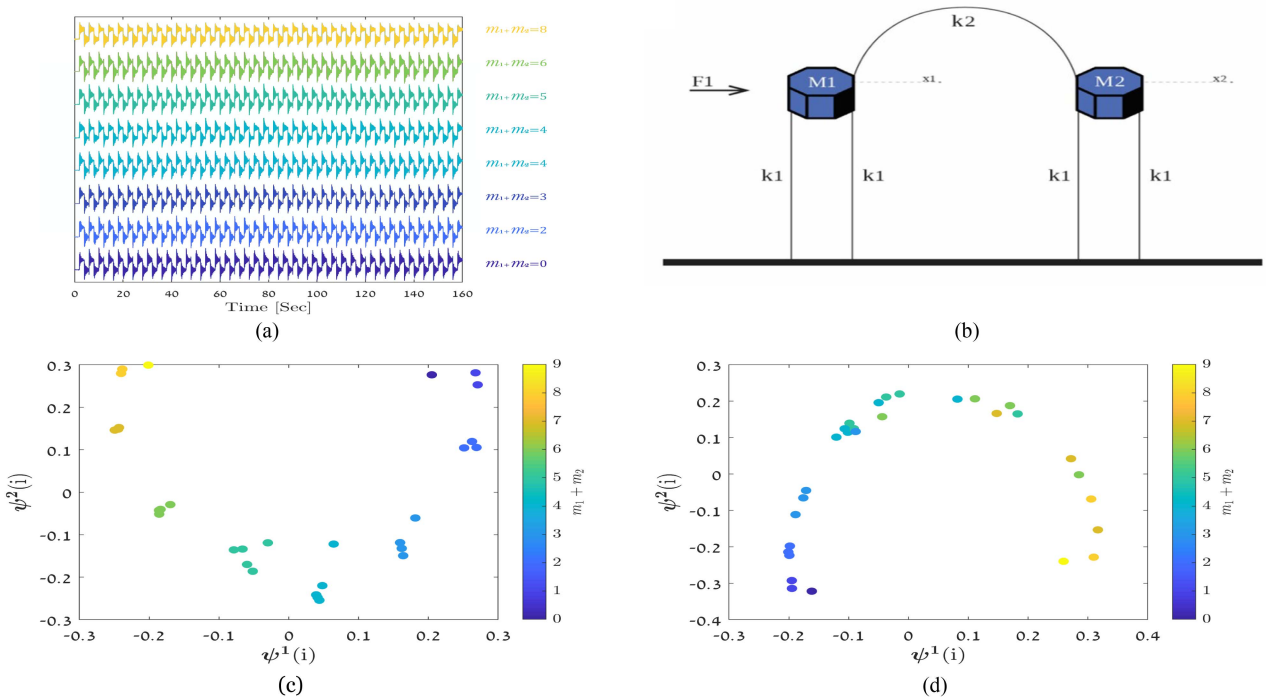

Fig. 2. Illustration of our method on a mechanical system. (a) An example of the input data to the algorithm: measurements of the location of mass m_2 over time. (b) A diagram of the mechanical system consisting of two coupled masses. (c) The two-dimensional embedding obtained by Algorithm 1 (our method) colored by the sum of the masses. (d) The two-dimensional embedding obtained by a modified Algorithm 1 colored by the sum of the masses, where the ℓ_2 norm is used instead of the Mahalanobis distance (see text for details).

$$= \frac{1}{2}(2k_1 + k_2) \cdot (m_1 + m_2)$$
$$\pm \frac{1}{2}(2k_1 + k_2) \cdot (m_1 + m_2)$$
$$\times \sqrt{1 + \frac{4 \cdot k_2^2 \cdot m_1 \cdot m_2}{(2k_1 + k_2)^2 \cdot (m_1 + m_2)^2}}$$

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 $m_1 + m_2$, since the term consisting of their product $m_1 \cdot m_2$ is of smaller order of magnitude. This implies that the hidden state variable in each trial could be parameterized by $\bar{\theta}_i = m_1 + m_2$.

2) Results: We apply Algorithm 1 to the input data $\{\boldsymbol{y}_i(t_j)\}_{j=1}^M$. As a baseline, we apply a similar algorithm to the same data, but instead of using the modified Mahalanobis distance we use the Euclidean distance between the features $\boldsymbol{z}_i$ (in Step 2 of Algorithm 1, the affinity matrix W is computed using $\|\boldsymbol{z}_i - \boldsymbol{z}_l\|_2^2$ instead of $d(\boldsymbol{z}_i, \boldsymbol{z}_l)$). Fig. 2(c)–(d) displays a two dimensional representation of the measurements resulting from the applications of the two algorithms. Fig. 2(c) shows the results of Algorithm 1 and Fig. 2(d) 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 two dimensional representation obtained by Algorithm 1 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 using the modified Mahalanobis distance rather than the Euclidean distance between the features $\boldsymbol{z}_i$ is critical for the recovery of the true hidden state of the system.

E. Theoretical Analysis of the Extraction of the State Variables

In this sub-section, we present the theoretical analysis supporting the proposed method. We recall that our goal is to find a pairwise distance between features of the measurements that satisfies (4), thereby revealing the distance between the hidden state variables. For this purpose, we compute a suitable set of features for each set of measurements that carries sufficient information about the state variables. Assuming that the process $\boldsymbol{y}_i(t)$ is stationary and ergodic and that we have an infinitely large number of measurements from each state, the two features in (5) and (6) can be recast as:

$$\boldsymbol{z}_i = \mathbb{E}[\boldsymbol{y}_i(t)] \tag{10}$$

$$C_i = \mathrm{Cov}[\boldsymbol{\mu}_i(t)] = \mathrm{Cov}[\boldsymbol{y}_i(t + \delta t)|\boldsymbol{y}_i(t)]. \tag{11}$$

Namely, the expected value of the Itô process measurements at a specific state, and the covariance matrix of the measurement increments. Then, as in (7), use a modified version of the Mahalanobis distance [20] between the features:

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) = \frac{1}{2}(\boldsymbol{z}_i - \boldsymbol{z}_l)^\top (C_i^{-1} + C_l^{-1})(\boldsymbol{z}_i - \boldsymbol{z}_l). \tag{12}$$

In the sequel, we show that this distance indeed reveals the distance between the hidden state variables.

Our analysis is divided into two cases as follows.

1) Direct Access: In order to simplify the exposition, we consider the case in which the measurements $\boldsymbol{y}_i(t)$ have direct access to system variables and are equal to $\boldsymbol{x}_i(t) = (\boldsymbol{\theta}_i(t), \boldsymbol{\eta}_i(t)) \in \mathbb{R}^d$.

Proposition 1: Given $\boldsymbol{x}_i(t) = (\boldsymbol{\theta}_i(t), \boldsymbol{\eta}_i(t)) \in \mathbb{R}^d$, the modified Mahalanobis distance in (12) between the features $\boldsymbol{z}_i = \mathbb{E}[\boldsymbol{x}_i(t)]$ using the covariance matrices $C_i = \text{Cov}[\boldsymbol{x}_i(t+\delta t)|\boldsymbol{x}_i(t)]$ can be written in terms of the Euclidean distance between the underlying state variables as follows:

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) = \frac{1}{\delta t}\left[\|\bar{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}}_l\|^2 + O(\epsilon)\right]. \tag{13}$$

See Appendix I for proof.

Proposition 1 implies that by assuming direct access to the state and noise variables, the modified Mahalanobis distance with the proposed features satisfies (4). Note that in this case, although we have access to the system variables $\boldsymbol{x}_i(t)$, we do not know which of them is a state variable $\boldsymbol{\theta}_i$ and which is considered noise. Our features and distance function enables us to separate the two kinds of variables and to obtain a distance between the state variables in an implicit manner without prior knowledge.

2) Non-Linear Measurements: We now consider the case where the observations are a function of the system variables, i.e., $\boldsymbol{y}_i(t) = f(\boldsymbol{x}_i(t))$, where $f : \mathbb{R}^d \to \mathbb{R}^s$ is some smooth bi-Lipschitz function. Considering this case requires an additional assumption on the function f and a small modification of the features $\boldsymbol{z}_i$, which are described next. Note that in Appendix III, we analyze a simpler case where f is linear.

Assumption 1: For any two realizations $\boldsymbol{x}_i$ and $\boldsymbol{x}_l$ of state and noise variables, we have:

$$\frac{(f^k_{p_1p_2}(\boldsymbol{x}_i) - f^k_{p_1p_2}(\boldsymbol{x}_l))^2}{f^k_{p_1}(\boldsymbol{x}_i)f^k_{p_2}(\boldsymbol{x}_i) + f^k_{p_1}(\boldsymbol{x}_l)f^k_{p_2}(\boldsymbol{x}_l)} \ll 1, \tag{14}$$

for $1 \le k \le s$ and $1 \le p_1, p_2 \le d$, where the subscripts correspond to partial derivatives, i.e., $f^k_p = \frac{\partial f^k}{\partial x^p}$, $f^k_{p_1p_2} = \frac{\partial^2 f^k}{\partial x^{p_1}\partial x^{p_2}}$, and the 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.

Assumption 1 could be viewed as an additional smoothness property of the observation function. This assumption holds for functions that have small local changes in their gradient, and it includes any second-order polynomial function.

Proposition 2: 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 1, the modified Mahalanobis distance with the features

$$\boldsymbol{z}_i = \mathbb{E}\left[\lim_{\delta t \to 0} \frac{\boldsymbol{y}_i(\delta t + t) - \boldsymbol{y}_i(t)}{\delta t} + \boldsymbol{y}_i(t)\right]$$
$$C_i = \text{Cov}[\boldsymbol{y}_i(t+\delta t)|\boldsymbol{y}_i(t)] \tag{15}$$

can be expressed as:

$$d(\boldsymbol{z}_i, \boldsymbol{z}_l) = \|\bar{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}}_l\|^2 + O(\|\boldsymbol{y}_i - \boldsymbol{y}_l\|^4) + O(\epsilon). \tag{16}$$

See Appendix II for proof.

Note that in the non-linear case, our theoretical analysis (see derivation in Appendix II) suggests using a feature $\boldsymbol{z}_i$ (15) that is different from the one suggested in (10). We show in the sequel that in practice when we estimate the features from the measurements, these two features coincide. Both Proposition 1 and Proposition 2 imply that, either with direct access or through some unknown observation function, the modified Mahalanobis distance between the proposed features reveals the distance between the state variables and attenuates the contribution of the noise variables. We conclude this subsection with a couple of remarks.

Remark 1: In the specific case where f is the identity function, i.e., $f(\boldsymbol{x}) = \boldsymbol{x}$, different features $\boldsymbol{z}_i$ can be computed: either (10) or (15)). Indeed, using the additional prior information of having direct access to the state and noise variables leads to a better approximation in Proposition 1 compared to Proposition 2. In the sequel, we will show that in practice the two definitions of $\boldsymbol{z}_i$ coincide.

Remark 2: Our method relies on the modified Mahalanobis distance proposed in [20]. In [19], this distance was used and analyzed in the context of temporal data stemming from a multi-scale dynamical system. The theoretical and algorithmic parts of the present work extend [20] and [19] in two main aspects. First, our work considers a different model than the models considered in [20] and in [19]. Here, the model consists of controlling variables, which are separated into (desired) intrinsic state variables and (undesired) noise variables. In addition, the model includes different dynamic regimes. By relying on the analysis in [19], we present new theoretical results, and consequently, derive new features that are specific for the model considered here. Importantly, in the sequel, we empirically support our model, features, and theoretical results by showing experimental results on DBS.

3) Feature Estimators: In order to estimate the features in (10) and (11) from the measurements at hand, we use the following assumptions. First, we assume that the number of measurements at each state M is large. Second, we assume that the measured signal $\boldsymbol{y}_i(t_j)$ is stationary and ergodic with respect to t_j at a fixed state i.

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

$$\hat{\boldsymbol{z}}_i = \frac{1}{M}\sum_{j=1}^{M} \boldsymbol{y}_i(t_j), \tag{17}$$

where the relation between the estimator and the desired feature is given by:

$$\boldsymbol{z}_i = \hat{\boldsymbol{z}}_i + O\left(\frac{1}{\delta t M}\right). \tag{18}$$

Since M is assumed to be large, $\hat{\boldsymbol{z}}_i$ is considered as a good approximation of $\boldsymbol{z}_i$.

The relation in (18) stems from the following derivation. Since 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:

$$\boldsymbol{z}_i = \mathbb{E}\left[\frac{1}{\delta t}(\boldsymbol{y}_i(\delta t + t) - \boldsymbol{y}_i(t)) + \boldsymbol{y}_i(t)\right]$$

$$= \frac{1}{M}\sum_{j=1}^{M}\left[\frac{\boldsymbol{y}_i(t_j)-\boldsymbol{y}_i(t_{j-1})}{\delta t}+\boldsymbol{y}_i(t_{j-1})\right]$$
$$= \frac{1}{M}\sum_{j=1}^{M}\left[\frac{\boldsymbol{y}_i(t_j)-\boldsymbol{y}_i(t_{j-1})}{\delta t}\right]+\frac{1}{M}\sum_{j=1}^{M}\boldsymbol{y}_i(t_{j-1})$$
$$= \frac{1}{M}\sum_{j=1}^{M}\left[\frac{\boldsymbol{y}_i(t_j)-\boldsymbol{y}_i(t_{j-1})}{\delta t}\right]+\hat{\boldsymbol{z}}_i. \quad (19)$$

The first term in (19) is a telescopic sum; after cancellation, this term is equal to $O(\frac{1}{\delta t M})$, leading to (18). We obtain that the feature estimate $\hat{\boldsymbol{z}}_i$ in (17) is a good estimate of the proposed feature (15) in the nonlinear case by (18). We note that the same feature estimate in (17) can serve as a good estimate of the proposed feature (10) in the direct access case as well. Therefore, in practice, we can use the same feature estimate for both cases.

III. Experimental Results on DBS

In this section, we show experimental results on the unsupervised detection of target regions during DBS surgery. We focus on two particular detection tasks: finding the subthalamic nucleus (STN) region and a sub-territory within the STN region, called the dorsolateral oscillatory region (DLOR).

A. Data and Preprocessing

The dataset was collected at the Hadassah University Medical Center, and it is completely anonymized without any personal identification information. Written informed consent was obtained from all patients and the study was approved by the Institutional Review Board of Hadassah Hospital in accordance with the Helsinki Declaration (reference code: 0168-10-HMO).

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. The data was acquired using a Neuro Omega system. The raw data was sampled at 44 kHz by a 16-bit A/D converter (using $\pm 1.25\,V$ input range; i.e., $\sim 2\,\mu V$ amplitude resolution). Then, the raw signal was bandpass filtered from 0.075 to 10 k Hz, using hardware of 2 and 3 pole Butterworth filters, respectively.

In the context of the problem setting described in Section II, we refer to each specific depth as a system state, denoted by i, for $i = 1, .., N$. Accordingly, let $g_i(\tau_j), j = 1 \ldots, T_i$ be the time series of the signal recorded at depth i, where τ_j is the discrete time index and T_i is the length of the signal recorded at depth i. We apply a pre-processing stage to the 1D time series measurements $g_i(\tau_j)$, where we compute the scattering transform [35], [36]. Scattering transform is based on a cascade of wavelet transforms and modulus operators and has been shown to yield an informative representation of time-series measurements. Let $\boldsymbol{y}_i(t_j)$ denote the resulting scattering transform at time $j = \{1\ldots, M_i\}$ and at depth $i = \{1\ldots, N\}$, 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 Fig. 3(a), where we plot the time series measurements from 6 different depths (states), colored according to the experts' 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 [13], [14], [15]. We also observe that there is no evident difference between the two classes of signals within the STN region, indicating that DLOR detection is a challenging task.

B. Detection of Sub-Territories of the Subthalamic Nucleus

1) Subthalamic Nucleus (STN) Detection: The proposed algorithm for the detection of the STN region appears in Algorithm 2, where we denote by EDT(i) the ith coordinate of the Estimate Distance from Target (EDT) vector designating the specific depth along the pre-planned trajectory. Note that the term EDT is frequently used in this context, and it is known in advance and does not suggest any a-priori knowledge of the target location.

Our empirical examination suggests that the STN location can be determined by the most dominant component, i.e., the entries of eigenvector $\boldsymbol{\psi}^1$ corresponding to the largest eigenvalue. Therefore, the detection of the STN is based only on $\boldsymbol{\psi}^1$ resulting from the application of Algorithm 1 with $P = 1$ 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 the entries of $\boldsymbol{\psi}^1$, namely, $\{\boldsymbol{\psi}^1(1), \boldsymbol{\psi}^1(2), \ldots\}$, indicating the entry to the STN region, where the indices of the vector entries represent the depth (state). In order to alleviate the effect of small perturbations, we smooth the sequence of entries of $\boldsymbol{\psi}^1$ by a moving average with a window of size 3. Then, we detect the transition by computing the difference between the medians at two consecutive running windows of size 5 and obtain $\tilde{\boldsymbol{\psi}}^1$. The index at which the maximal difference is obtained is denoted by i_{en}, indicating the depth of the entry point to the STN EDT(i_{en}). Once the entry point is determined, the exit point is set as the first point at which $\boldsymbol{\psi}^1(i)$ is smaller than $\boldsymbol{\psi}^1(i_{\text{en}})$. We denote the index of the exit point by i_{ex} and the corresponding depth by EDT(i_{ex}).

The result of the application of Algorithm 2 to the example presented in Fig. 3(a) is shown in Fig. 3(b)–(c), where we plot $\boldsymbol{\psi}^1(i)$ and $\tilde{\boldsymbol{\psi}}^1(i)$ 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 $\boldsymbol{\psi}^1(i)$ by sign$(\delta)\cdot\boldsymbol{\psi}^1(i)$, where

$$\delta = |\max_{i=2,\ldots,N}(\tilde{\boldsymbol{\psi}}^1(i)) - \tilde{\boldsymbol{\psi}}^1(1)| - |\min_{i=2,\ldots,N}(\tilde{\boldsymbol{\psi}}^1(i)) - \tilde{\boldsymbol{\psi}}^1(1)|.$$

2) Dorsolateral Oscillatory Region (DLOR) Detection: The proposed algorithm for the detection of the DLOR appears in Algorithm 3. Since the transitions between the sub-territories

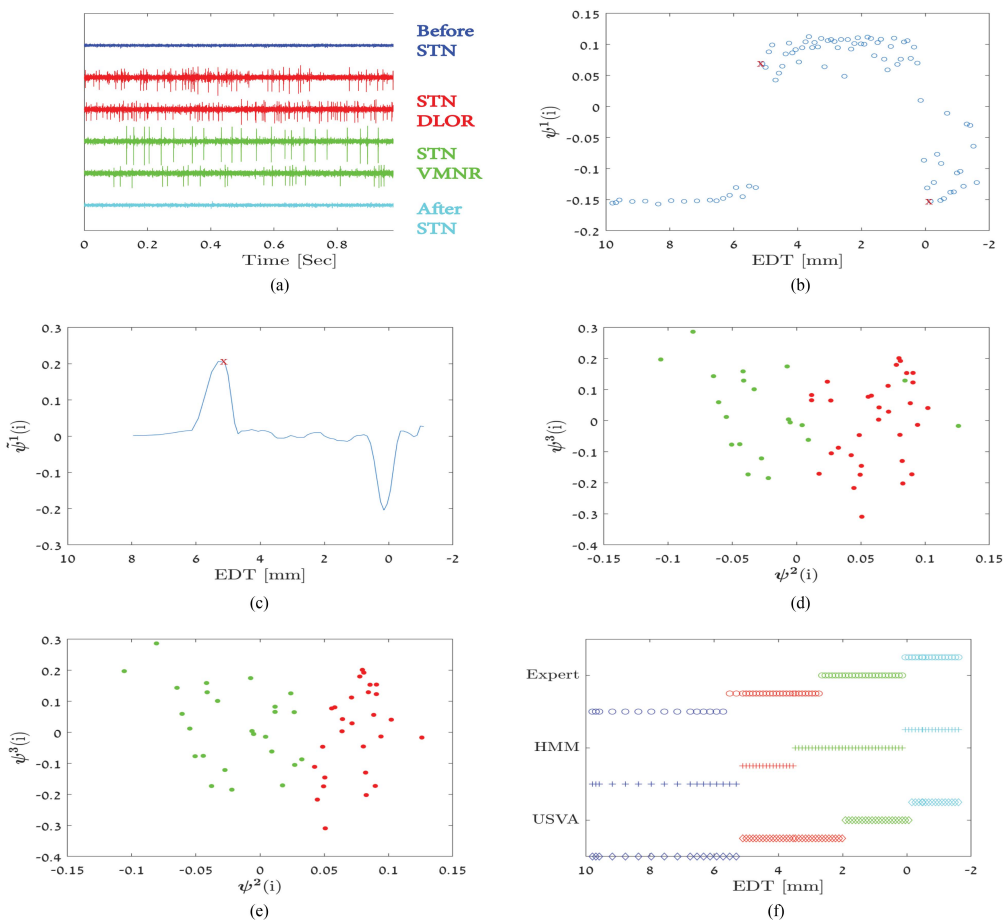

Fig. 3. STN border detection obtained by our proposed method - Unsupervised State Variables approximation (USVA) - applied to a single trajectory. (a) The input data – time series of measurements of the neuronal activity along the pre-planned trajectory. We show 6 representative raw signal traces out of the 84 signal traces at various depths along the trajectory (a single DBS track) recorded from a Parkinson's disease patient. The different time series are colored according to the experts' labels: white matter before STN in blue, Dorso-Lateral Oscillatory Region (DLOR) in red, Ventro Medial Non-oscillatory Region (VMNR) in green, and white matter after STN in light blue. (b) The 1D embedding obtained by the USVA method. The Y-axis displays the approximate state variable value, i.e., the value of the most dominant eigenvector as a function of the Estimated Distance from Target (EDT). The STN entry and exit locations marked by a human expert are marked by red 'x'. (c) Same as (b) but after pre-processing (smoothing) with a moving average window of size 3, and computing the difference between the medians of consecutive windows of size 5. (d) 2D embedding obtained by the USVA method. The embedded points are colored according to the expert's labels, where red points belong to depths in the DLOR and green points belong to depths in the VMNR. The X-axis indicates the value of the second most dominant eigenvector and the Y-axis indicates the corresponding value of the third most dominant eigenvector. (e) Same 2D embedding as in (d) but colored according to k-means as suggested in Algorithm 3. (f) A comparison between the detection results obtained by our USVA method (marked with '◇'), HMM (marked with '+'), and the expert's labels (marked with 'o'). Colors are the same as in Fig. 3(a).

of the STN region are subtle and not as distinct as the transition into and out of the STN, we perform two adjustments with respect to Algorithm 2. First, we assume that the information on subtle changes in the system's state variables is manifested deeper in the spectrum, namely in eigenvectors corresponding to 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:

$$K^t = K + K^s \tag{20}$$

where K is the operator defined in (9) and is used for the detection of the STN region, and K^s is a kernel that enhances temporal proximity, which is given by:

$$K^s_{i,l} = \frac{W^s_{i,l}}{\boldsymbol{w}^s(i)} \quad , \quad \boldsymbol{w}^s(i) = \sum_{l=1}^{N} W^s_{i,l} \tag{21}$$

where

$$W^s_{i,l} = \exp\left\{-\frac{\|\mathrm{EDT}(i) - \mathrm{EDT}(j)\|^2}{\epsilon_s}\right\}, \tag{22}$$

where ϵ_s is the kernel scale.

In accordance with the above adjustments, we apply eigenvalue decomposition to K^t and represent each depth in the STN region using *two* eigenvectors, ψ^2 and ψ^3, corresponding to the third and fourth largest eigenvalues. We exclude ψ^0 and ψ^1 because ψ^0 is trivial and ψ^1 contains information on the STN boundaries, which we already exploited in Algorithm 2. We note

Algorithm 2: 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 $\in \mathbb{R}^N$ – Vector indicating the Estimated Distance from Target of each depth.

Output: *STN entry point* and *STN exit point*.

1) For each time series $g_i(\tau_j)$, compute its Scattering Transform: $\boldsymbol{y}_i(t_j) = \Phi(g_i(\tau_j)) \in \mathbb{R}^k$, where $j = 1, \ldots, M_i$, $i = 1, .., N$, and Φ represents the Scattering transform operator
2) Compute $\boldsymbol{\psi}^1$ according to Algorithm 1
3) Compute $\tilde{\boldsymbol{\psi}}^1$ by applying a moving average with a window of size 3 samples to $\boldsymbol{\psi}^1$, and then, compute the difference between the medians at two consecutive running windows of size 5 samples
4) $i_{\text{en}} = \underset{i}{\text{argmax}} \quad \tilde{\boldsymbol{\psi}}^1(i)$
5) $i_{\text{ex}} = \underset{i}{\text{argmin}} \quad \{i : \boldsymbol{\psi}^1(i) \leq \boldsymbol{\psi}^1(i_{\text{en}}) \quad \text{and} \quad i > i_{\text{en}}\}$
6) Set the *STN entry point* as EDT(i_{en}), and the *STN exit point* as EDT(i_{ex})

Algorithm 3: DLOR Detection.

Input: The affinity matrix $K \in \mathbb{R}^{N \times N}$, the STN entry point, the STN exit point
EDT $\in \mathbb{R}^N$ – the vector indicating the Estimated Distance from Target of each depth.

Output: The *DLOR exit point*.

1) Compute a smoothing kernel K^s according to (21)
2) Compute the kernel K^t according to (20)
3) Apply eigenvalue decomposition to K^t and obtain its eigenvalues and eigenvectors
4) Represent each depth in the STN region according to (23), i.e., by $R(i) = (\boldsymbol{\psi}^2(i), \boldsymbol{\psi}^3(i), \text{EDT}(i))$
5) Divide all depth representations $R(i)$ into 2 clusters, DLOR and STN-VMNR, using K-means initialized with the STN entry (i_{en}) and exit (i_{ex}) points
6) Set $i_d = \underset{i}{\text{argmin}}\{R(i) \in \text{STN} \quad \text{-VMNR}\}$
7) Set the *DLOR exit point* as EDT(i_d)

that we examined the use of different numbers of eigenvectors, and we choose to represent each depth using two eigenvectors since it led to the best empirical results. Yet, our empirical test suggests that the results are not sensitive to using a different number of eigenvectors.

This representation enables us to find separation in the embedded space. Specifically, we apply K-means [37] to the following embedding of each depth within the STN region:

$$R(i) = (\boldsymbol{\psi}^2(i), \boldsymbol{\psi}^3(i), \text{EDT}(i)). \tag{23}$$

We note that since the DLOR is a continuous region we added the third coordinate that encourages temporal continuity of the separation. This third coordinate is appropriately scaled so that it fits the dynamical range of the other two coordinates. We apply K-means with $k = 2$, initialized with the entry depth and exit depth of the STN region, and obtain two clusters within the STN. The cluster that includes the entry depth to the STN region is denoted as the DLOR, and the other cluster, which includes the exit depth from the STN region, is denoted as the STN-VMNR. An example of the 2D embedding of measurements from depths within the STN region is shown in Fig. 3(d)–(e). In Fig. 3(d), the points are colored according to the 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 Fig. 3(e). Indeed, we see that our unsupervised detection coincides with the expert's labels. We note that although the embedding in Fig. 3(d) may suggest that the classification of the DLOR/STN-VMNR could be based on the sign of ψ^2, inspecting other trajectories indicates that it does not apply in general and that the clustering should be based on a combination of ψ^2 and ψ^3.

Finally, based on Algorithms 2 and 3, we cluster the data according to the 4 labels. A visual comparison between the labels obtained by our unsupervised method and by the supervised HMM algorithm with respect to labels given by an expert to data from a specific example is presented in Fig. 3(f). For convenience, our method is denoted by USVA (Unsupervised State Variables Approximation). We see that in the presented example, our unsupervised method is able to detect the STN region with an accuracy that is comparable to the accuracy of the supervised HMM method. In addition, our unsupervised method obtains a superior detection of the DLOR compared to the supervised HMM. We note that these results are with respect to the expert's labels.

C. Quantitative Detection Results

We apply our method (Algorithm 2 and Algorithm 3) to 25 different trajectories recorded from 16 patients, and we compare the results to the results obtained by the HMM algorithm proposed in [13], which is considered the gold-standard. Each trajectory consists of three transition points of interest: the STN entry, the DLOR exit, and the STN exit (note that the DLOR entry is the same as the STN entry). In order to quantitatively evaluate the detection, we measure 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 entry and exit errors (divided by the size of the STN region), DLOR entry and exit errors (divided by the size of the DLOR region), and the overall STN and DLOR errors. Note that the STN entry error and the DLOR entry error differ only by the normalization factor, since the STN entry point coincides with the DLOR entry point. The median and interquartile range (IQR) of the five performance measures are reported in Fig. 4. To complement the experimental study, we also report their mean and standard deviation in percentage in Table I and Table II. We remark that in our performance evaluation, failures to detect the DLOR exit point is considered to be a 100% error. In addition, we note that

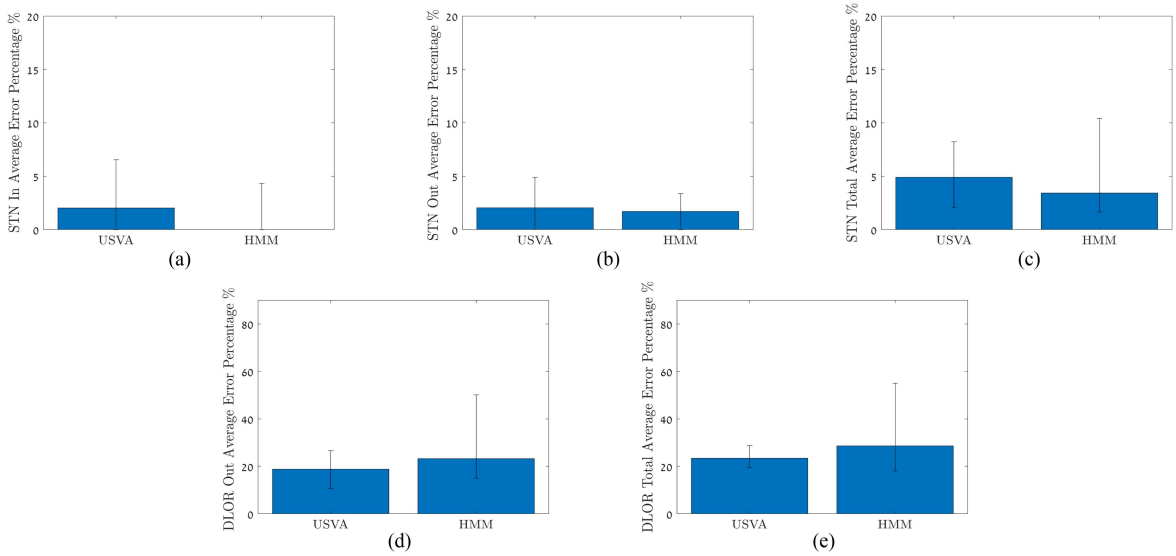

Fig. 4. Comparison between our method and the HMM algorithm. The comparison is based on the experts' labels of 25 different trajectories of 16 different patients. The blue thick bars indicate the median error percentage of each task and the black thin bars indicate the interquartile range (IQR). (a) The median error in detecting the entry point to the STN region. (b) The median error in detecting the exit from the STN region. (c) The median overall error in detecting the STN region. (d) The median error in detecting the exit from the DLOR. (e) The median overall error in detecting the DLOR.

TABLE I
THE STN BORDERS DETECTION RESULTS OBTAINED BY OUR METHOD(USVA) AND THE HMM ALGORITHM. THE PRESENTED VALUES ARE THE AVERAGE BORDER DETECTION ERROR (OVER THE 25 TESTED TRAJECTORIES) AND THE STANDARD DEVIATION WITH RESPECT TO THE IDENTIFICATION OF A HUMAN EXPERT. THE ERROR IS REPORTED IN PERCENTAGE RELATIVE TO THE STN SIZE

	STN entry	STN exit	STN overall
USVA	4.77 ± 6.97	3.95 ± 5.53	8.72 ± 11.05
HMM	4.41 ± 8.62	2.89 ± 4.56	7.31 ± 9.8

TABLE II
SAME AS IN TABLE I, BUT FOR THE DLOR DETECTION. THE ERROR IS REPORTED IN PERCENTAGE RELATIVE TO THE DLOR SIZE

	DLOR exit	DLOR overall
USVA	17.89 ± 13.68	24.49 ± 17.41
HMM	39.16 ± 34.87	43.86 ± 37.03

the DLOR performance measure values are higher than the STN performance measures; this is due to the normalization by the size of the DLOR region, which is smaller than the STN region.

We observe that our method attains results comparable to the gold-standard in the detection of the STN and outperforms it in the detection of the DLOR. Importantly, our method is unsupervised whereas the HMM-based method is supervised, and therefore, could be biased towards the labeling of the specific expert labels, which were used for training.

D. Run Time

The run-time of our algorithm is mainly governed by the pre-processing stage of each MER (scattering transform). Therefore, it highly depends on the number of MERs taken from different trajectory depths and the recording time of each MER. After acquiring the MERs from the entire trajectory, the total run-time of our method is approximately 3-5 minutes on a standard personal computer. This run-time could be significantly reduced to a few seconds if we apply the pre-processing stage during the recording of each MER.

IV. CONCLUSION

We presented a method that can enable physicians with accurate detection of the STN and the DLOR during DBS surgery. The MERs collected during surgery can be fed into a system that implements Algorithms 2 and 3. This system, which does not need to be manually tuned, provides recommendations regarding the STN and the DLOR locations. By the nature of our method, these recommendations are not biased toward specific experts, and therefore, can help physicians validate their decisions. We note that our method recommends a location of the target regions only after all the MERs have been acquired.

Future work will address the detection of the Globus Pallidus (GP), which is an area of interest during a DBS surgery for treating advanced Parkinson's disease and dystonia [38]. Since the setup of the GP detection is very similar to the STN and DLOR detections, our unsupervised method can be applied with only mild adjustments. Another research direction concerns the estimation of the covariance of the features. Currently, we use the sample covariance, however, finding better estimators of the covariance matrix, for example using shrinkage [39], [40], may significantly improve the results. The generalization of the proposed method to multiple sets of measurements from different modalities could also be considered. In the context of manifold learning, multimodal data fusion has attracted much attention recently, e.g., [41], [42]. The proposed variant of the Mahalanobis distance could be incorporated into such multimodal

methods, facilitating unsupervised target and anomaly detection for multi-channel and multi-modal data. Finally, we remark that our algorithm generates an embedding of measurements from which the STN and the DLOR locations are identified. However, this embedding does not have an inverse map. In the context of DBS, if such a map existed, it could be used to extract the signal properties characterizing the different regions and the transitions between them. Future work will explore the latent governing variables of the MERs that determine the locations of the STN and the DLOR.

Acknowledgment

We would like to convey our gratitude to Hadassah University Medical Central for its willingness to share the data with us. We also wish to thank Omer Naor from Alpha Omega for fruitful early-stage discussions, and Or Yair, Pavel Lifshits, and Izhak Bucher for their help with the mechanical system experiment. We thank the Editor and the anonymous reviewers for their insightful comments and suggestions. Their review helped us to improve this manuscript significantly.

References

[1] P. Limousin et al., "Electrical stimulation of the subthalamic nucleus in advanced Parkinson's disease," *New England J. Med.*, vol. 339, no. 16, pp. 1105–1111, 1998.

[2] A. Benabid et al., "Acute and long-term effects of subthalamic nucleus stimulation in Parkinson's disease," *Stereotactic Funct. Neurosurg.*, vol. 62, no. 1–4, pp. 76–84, 1994.

[3] M. I. Hariz, "Complications of deep brain stimulation surgery," *Movement Disord.: Official J. Movement Disord. Soc.*, vol. 17, no. S3, pp. S162–S166, 2002.

[4] R. C. Nickl et al., "Rescuing suboptimal outcomes of subthalamic deep brain stimulation in Parkinson disease by surgical lead revision," *Neurosurgery*, vol. 85, no. 2, pp. E314–E321, 2019.

[5] E. Moro et al., "The impact on Parkinson's disease of electrical parameter settings in STN stimulation," *Neurology*, vol. 59, no. 5, pp. 706–713, 2002.

[6] K. Witt et al., "Factors associated with neuropsychiatric side effects after STN-DBS in Parkinson's disease," *Parkinsonism Related Disord.*, vol. 18, pp. S168–S170, 2012.

[7] A. Moran et al., "Real-time refinement of subthalamic nucleus targeting using Bayesian decision-making on the root mean square measure," *Movement Disord.: Official J. Movement Disord. Soc.*, vol. 21, no. 9, pp. 1425–1431, 2006.

[8] M. Weinberger et al., "Beta oscillatory activity in the subthalamic nucleus and its relation to dopaminergic response in Parkinson's disease," *J. Neuriophysiol.*, vol. 96, no. 6, pp. 3248–3256, 2006.

[9] A. Zaidel et al., "Subthalamic span of β oscillations predicts deep brain stimulation efficacy for patients with Parkinson's disease," *Brain*, vol. 133, no. 7, pp. 2007–2021, 2010.

[10] R. R. Shamir et al., "Microelectrode recording duration and spatial density constraints for automatic targeting of the subthalamic nucleus," *Stereotactic Funct. Neurosurg.*, vol. 90, no. 5, pp. 325–334, 2012.

[11] P. Novak et al., "Detection of the subthalamic nucleus in microelectrographic recordings in Parkinson disease using the high-frequency (>500 Hz) neuronal background," *J. Neurosurg.*, vol. 106, no. 1, pp. 175–179, 2007.

[12] I. Telkes et al., "Prediction of STN-DBS electrode implantation track in Parkinson's disease by using local field potentials," *Front. Neurosci.*, vol. 10, 2016, Art. no. 198.

[13] D. Valsky et al., "Stop! border ahead: Automatic detection of subthalamic exit during deep brain stimulation surgery," *Movement Disord.*, vol. 32, no. 1, pp. 70–79, 2017.

[14] S. Wong et al., "Functional localization and visualization of the subthalamic nucleus from microelectrode recordings acquired during DBS surgery with unsupervised machine learning," *J. Neural Eng.*, vol. 6, no. 2, 2009, Art. no. 026006.

[15] A. Zaidel et al., "Delimiting subterritories of the human subthalamic nucleus by means of microelectrode recordings and a hidden Markov model," *Movement Disord.*, vol. 24, no. 12, pp. 1785–1793, 2009.

[16] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, 2006.

[17] J. C. Mattingly et al., "Convergence of numerical time-averaging and stationary measures via poisson equations," *SIAM J. Numer. Anal.*, vol. 48, no. 2, pp. 552–577, 2010.

[18] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Berlin, Germany: Springer, 2012.

[19] C. J. Dsilva et al., "Data-driven reduction for a class of multiscale fast-slow stochastic dynamical systems," *SIAM J. Appl. Dyn. Syst.*, vol. 15, no. 3, pp. 1327–1351, 2016.

[20] A. Singer and R. R. Coifman, "Non-linear independent component analysis with diffusion maps," *Appl. Comput. Harmonic Anal.*, vol. 25, no. 2, pp. 226–239, 2008.

[21] D. Kushnir et al., "Anisotropic diffusion on sub-manifolds with application to earth structure classification," *Appl. Comput. Harmon. Anal.*, vol. 32, no. 2, pp. 280–294, 2012.

[22] J. B. Tenenbaum et al., "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 260, pp. 2319–2323, 2000.

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

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

[25] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural. Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[26] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, 2006.

[27] R. Talmon et al., "Manifold learning for latent variable inference in dynamical systems," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3843–3856, Aug. 2015.

[28] O. Yair et al., "Reconstruction of normal forms by learning informed observation geometries from data," *Proc. Nat. Acad. Sci.*, vol. 114, no. 38, pp. E7865–E7874, 2017.

[29] B. Zhu et al., "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[30] A. Singer et al., "Diffusion interpretation of nonlocal neighborhood filters for signal denoising," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 118–139, 2009.

[31] R. Ibanez et al., "A manifold learning approach to data-driven computational elasticity and inelasticity," *Arch. Comput. Methods Eng.*, vol. 25, no. 1, pp. 47–57, 2018.

[32] T. Shnitzer et al., "Recovering hidden components in multimodal data with composite diffusion operators," *SIAM J. Math. Data Sci.*, vol. 1, no. 3, pp. 588–616, 2019.

[33] H.-T. Wu et al., "Assess sleep stage by modern signal processing techniques," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1159–1168, Apr. 2015.

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

[35] S. Mallat, "Group invariant scattering," *Commun. Pure Appl. Math.*, vol. 65, no. 10, pp. 1331–1398, 2012.

[36] J. Bruna et al., "Intermittent process analysis with scattering moments," *Ann. Statist.*, vol. 43, no. 1, pp. 323–351, 2015.

[37] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *J. Roy. Stat. Soc. Ser. C (Appl. Statist.)*, vol. 28, no. 1, pp. 100–108, 1979.

[38] D. Valsky et al., "Real-time machine learning classification of pallidal borders during deep brain stimulation surgery," *J. Neural Eng.*, vol. 17, no. 1, 2020, Art. no. 016021.

[39] M. Gavish et al., "Optimal recovery of precision matrix for Mahalanobis distance from high-dimensional noisy observations in manifold learning," *Inf. Inference: J. IMA*, vol. 08, 2022, Art. no. iaac010.

[40] D. L. Donoho et al., "Optimal shrinkage of eigenvalues in the spiked covariance model," *Ann. Statist.*, vol. 46, no. 4, 2018, Art. no. 1742.

[41] R. R. Lederman and R. Talmon, "Learning the geometry of common latent variables using alternating-diffusion," *Appl. Comput. Harmon. Anal.*, vol. 44, no. 3, pp. 509–536, 2018.

[42] M. Salhov et al., "Multi-view kernel consensus for data analysis," *Appl. Comput. Harmon. Anal.*, vol. 49, no. 1, pp. 208–228, 2020.