ALTERNATING DIFFUSION MAPS FOR DEMENTIA SEVERITY ASSESSMENT

Tal Shnitzer[1], Maya Rapaport[1], Noga Cohen[1], Natalya Yarovinsky[2], Ronen Talmon[1] and Judith Aharon-Peretz[2]

[1] Viterbi Faculty of Electrical Engineering
Technion, Israel Institute of Technology
Haifa, 3200003, Israel

[2] Rambam Health Care Campus
Technion, Israel Institute of Technology
Haifa, 3200003, Israel

ABSTRACT

In this paper we address the detection of Alzheimer's disease based solely on EEG recordings. We assume that the state of Alzheimer's disease can be described by a latent manifold, captured by the EEG sensors and apply alternating diffusion to reveal this common underlying manifold from multiple EEG sensors. We show that based on a small number of EEG electrodes, a new representation can be obtained, which allows a clear distinction between healthy subjects and Alzheimer patients in different disease stages.

***Index Terms*—** Manifold Learning, Alternating Diffusion, Electroencephalogram, Alzheimer's Disease

1. INTRODUCTION

Alzheimer's disease, the most prevalent form of dementia [1], is a progressive neurodegenerative disorder [1] that becomes more prevalent with increasing age. Clinical manifestations include decline in cognitive and behavioral functions like memory, thinking and language skills. Alzheimer's disease has a long preclinical phase and a disease duration of 8-10 years [2]. Alzheimer's disease is commonly diagnosed according to core clinical manifestations and assessing their change with time [3]. At the time, Alzheimer's pathological changes could not be measured in vivo, so disease could be definitively diagnosed only after death. With the advent of cerebrospinal fluid (CSF) and imaging biomarkers, it is now possible to diagnose the Alzheimer's disease pathophysiological processes early in the course of the disease when there is still no evidence of cognitive impairment. Reliable biomarkers for disease progression are still lacking.

Electroencephalogram (EEG) was shown to be a reliable tool in the research and diagnosis of dementia, since frequency content and EEG complexity change as the disease progresses, due to neuronal connectivity decline [4]. Numerous studies have previously addressed the classification of Alzheimer's disease stage based on EEG recordings [5–8]. For example, in [6], an extensive review of different feature extraction methods and classification algorithms is presented. The authors show that a classification success rate of 0.88 can be achieved. However, a reliable classification of Alzheimer's disease state based on EEG recordings remains a challenge due to the significant artifacts and noise introduced by the EEG sensors.

In this paper, we analyze EEG recordings and assume that the state of Alzheimer's disease can be described by a latent manifold which is captured by the EEG sensors. Each sensor captures a deformed manifold due to the position of the electrodes, subject movement, measurement noise and non-linearities related to the cranium [9]. Therefore, we are interested in revealing the true underlying manifold, common to different EEG sensors. For this purpose, we apply alternating diffusion [10,11] to several sensors of EEG recordings. We demonstrate the ability of alternating diffusion to extract meaningful variables in a real application involving several sensors.

The paper is organized as follows. In Section 2 we present the basic concepts of diffusion maps [12], which lay the foundation to alternating diffusion. In Section 3, the problem formulation and alternating diffusion framework are presented. Finally, in Section 4, experimental results of the application of alternating diffusion to EEG recordings of Alzheimer patients and control subjects are presented.

2. BACKGROUND

Consider a data set of samples, e.g., data measured by a single EEG channel. Based on the recorded data, we wish to construct a low dimensional representation which describes the intrinsic properties of the system. We assume that the data approximately lie on a low dimensional Riemannian manifold $\mathcal{M}$, which is embedded in a higher dimensional ambient space. This manifold describes the latent state of the system. The data are samples on $\mathcal{M}$ with a density function q. Based on these data, we construct a new embedding which represents the underlying manifold by applying diffusion maps [12]. Following is a brief description of the setting and algorithm of diffusion maps.

Given a discrete set of samples $\{\boldsymbol{x}_i\}_{i=1}^N, \boldsymbol{x}_i \in \mathbb{R}^M$, we construct a pairwise affinity kernel $k_\epsilon(\boldsymbol{x}_i, \boldsymbol{x}_j)$ according to

$$k_\epsilon(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left\{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\epsilon}\right\} \tag{1}$$

where $\|\cdot\|^2$ denotes the Euclidean norm and $\epsilon > 0$ is the scale parameter which induces a notion of locality.

The kernel in (1) forms a weighted graph G in which the vertices are the data samples and the kernel defines the weights of the edges, i.e., the weight of the edge connecting the vertex $\boldsymbol{x}_i$ and the vertex $\boldsymbol{x}_j$ is $k_\epsilon(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

The kernel is then normalized by $d_\epsilon(\boldsymbol{x}_j) = \sum_{i=1}^N k_\epsilon(\boldsymbol{x}_i, \boldsymbol{x}_j)$ as follows

$$p_\epsilon(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{k_\epsilon(\boldsymbol{x}_i, \boldsymbol{x}_j)}{d_\epsilon(\boldsymbol{x}_j)} \tag{2}$$

The resulting normalized matrix $P_{i,j} = p_\epsilon(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is column stochastic and can be interpreted as the transposed transition probability matrix of a Markov chain on the graph vertices [13].

This work was supported by the Israel Science Foundation under Grant 1490/16.

Table 1: Subject classification details by device.

Device	No. of Subjects				
	Control	Mild	Moderate	Severe	Total
1	6	2	1	4	13
2	5	2	4	2	13
MMSE score	30/30	20-28/30	15-19/30	0-15/30	

Table 2: Pre-processing parameters for each device

Device	Sampling rate [Hz]	BPF Passband [Hz]	EEG Channels used for AD	Scattering Transform Window size [sec]	Total Size per Subject [sec]
1	512	[5,80]	Fpz,Fz,Cz	1	40
2	500	[2,80]	P7,P8	1.024	40

For sufficiently large ϵ, the graph is well connected, and the matrix $\mathbf{P}$ has an eigenvalue decomposition in which the eigenvalues, when denoted in descending order, satisfy $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1} \geq 0$. Based on the left eigenvectors of $\mathbf{P}$, denoted by $\boldsymbol{\phi}_n$, corresponding to the ℓ largest non-trivial eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_\ell$, we define an embedding of the given data set into an Euclidean space in $\mathbb{R}^\ell$:

$$\Phi(\boldsymbol{x}_i) = [\phi_1(i), ..., \phi_\ell(i)] \tag{3}$$

where ℓ is the dimensionality of the new embedded space.

It was shown in [14] that in the limit $N \to \infty, \epsilon \to 0$ the discrete Markov chain converges to a continuous diffusion process. Furthermore, the discrete matrix $(\mathbf{I} - \mathbf{P})/\epsilon$, where $\mathbf{I}$ is the identity matrix, converges to the following diffusion operator on the manifold $\mathcal{M}$

$$\lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{\mathbf{I} - \mathbf{P}}{\epsilon} = \Delta - \frac{\Delta q}{q} \tag{4}$$

where Δ is the Laplace operator and q is the density of the points on $\mathcal{M}$. Note that when the density function q is constant (uniform density), this operator approximates the Laplace-Beltrami operator on $\mathcal{M}$ [12]. Therefore, the embedding defined in (3) can be associated with the spectral decomposition of the Laplace-Beltrami operator on the manifold $\mathcal{M}$.

Recently, several studies have applied such manifold learning methods to analyze medical data, e.g. EEG recordings, heart rate measurements and respiratory sensors [15–20]. For example, in [15], alternating diffusion is applied to multimodal sleep recordings which include an airflow sensor and an abdominal motion sensor. In this paper, the authors showed that by applying alternating diffusion, the common source of variability in these sensors is extracted and that it captures the sleep stage information. In another work [20], the authors apply weighted locally linear embedding (WLLE) to EEG recordings in order to detect epileptic seizures. The WLLE method is used to obtain intrinsic features which are then classified using support vector machine (SVM).

3. PROPOSED ALGORITHM

Consider M EEG sensors measuring electrical brain activity. Our basic assumption is that the EEG sensors measure similar relevant information. Yet, EEG signals are known to be very noisy. To encode these two properties, we assume that the signal samples from the M sensors share a hidden underlying geometric structure (manifold), whereas each sensor introduces additional sensor-specific interferences and noise.

Let $x_{k,i}$ denote N signal samples from the k-th EEG sensor, where $i = 1, \ldots, N$ and $k = 1, \ldots, M$. We assume that $x_{k,i}$ are samples from a (hidden) composition of spaces $\mathcal{M}_k \times \mathcal{N}_k$, where $\mathcal{N}_k$ represents the space of sensor-specific interferences and noise, and $\mathcal{M}_k = g_k(\mathcal{M})$ is the k-th sensor view of a common manifold $\mathcal{M}$ through some arbitrary unknown function g_k.

In the context of this application, the common underlying manifold $\mathcal{M}$ represents intrinsic properties of the examined system, specifically, the severity of Alzheimer's disease. In contrast, each sensor-specific space $\mathcal{N}_k$ captures various artifacts and measurement noise specific to the sensor.

Our goal is therefore to infer a parametrization of the common manifold in a data-driven manner from the signal samples $x_{k,i}$, without assuming any additional rigid model assumptions. By accomplishing this, we show that the obtained parametrization is in good agreement with the severity and progression of Alzheimer's disease.

For this purpose, we apply alternating diffusion [10, 11, 15] to the M sensors. The alternating diffusion method is described in the remainder of this section. We note that in this work, we consider more than two sensors due to the significant noise introduced by the EEG modality.

Given data recorded by M sensors, diffusion maps is applied to each sensor, resulting in M operators $\{\mathbf{P}_k\}_{k=1}^M$ calculated according to (2). The alternating diffusion operator is then defined by

$$\mathbf{A} = \prod_{k=1}^{M} \mathbf{P}_k \tag{5}$$

Note that $\mathbf{A}$ is a column-stochastic matrix. As a result, it can be seen as a transition probability matrix of a Markov chain, where each step consists of M steps corresponding to $\mathbf{P}_k$. Several consecutive steps of this Markov chain are alternating steps of $\mathbf{P}_k$. Intuitively, in each step k, the application of the operator $\mathbf{P}_k$ can be interpreted as a propagation of the diffusion from each point to the highly connected points to it according to the graph structure obtained by sensor k. In the consecutive step, $\mathbf{P}_{k+1}$ is applied which results in a propagation from each point to the highly connected points to it based on sensor $k+1$. Combining two consecutive steps results in an "effective" propagation from each point to points which are highly connected to it in *both* sensors k and $k+1$. Therefore, in the resulting operator

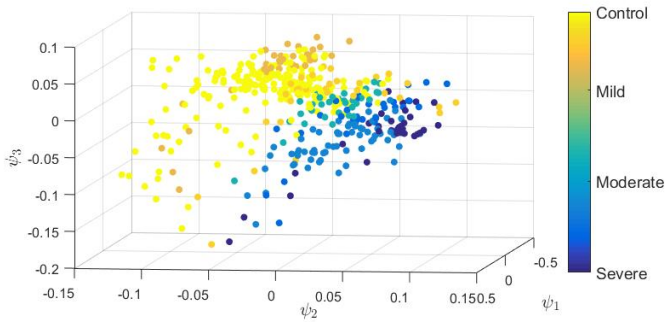


Fig. 1: Alternating diffusion coordinates of the first EEG device. The points are colored based on the MMSE score of each subject.

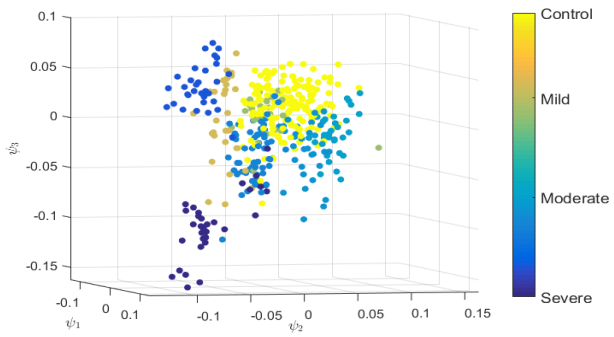


Fig. 2: Alternating diffusion coordinates of the second EEG device. The points are colored based on the MMSE score of each subject.

$\mathbf{A}$ (comprising of M steps), a diffusion propagation between one point to another is attained only if these points are connected in the graph structures obtained by *all* sensors. As we will indicate in the sequel, preserving connections obtained by all sensors enhances the common structure underlying all sensors, as well as attenuating the sensor-specific variables.

Based on the Eigenvalue Decomposition of $\mathbf{A}$, we define an embedding of the given data sets into an Euclidean space, similarly to (3); the embedding is constructed by taking the first ℓ left eigenvectors, $\boldsymbol{\psi}_n$, of $\mathbf{A}$, corresponding to the largest ℓ eigenvalues:

$$\Psi\left(\boldsymbol{x}_i\right) = \left[\psi_1\left(i\right), \ldots \psi_\ell\left(i\right)\right] \tag{6}$$

In [11] it was shown that in the limit $N \to \infty, \epsilon \to 0$, $\mathbf{A}$ converges to a continuous operator defined on the common manifold $\mathcal{M}$, similarly to (4). As a result, the alternating diffusion embedding (6) is equivalent to the diffusion maps embedding (3) we would have obtained if we had direct access to samples from the hidden common manifold $\mathcal{M}$. This implies that the embedding defined in (6) can be seen as a parametrization of the common hidden manifold $\mathcal{M}$, obtained from deformed and noisy samples from multiple sensors.

Note that the construction of the embedding described in this section is essentially feature extraction which provides a new feature space, representing the EEG data from multiple sensors. The obtained features can be used as an input for classification algorithms.

To summarize, the main challenges in the analysis of EEG data are the significant noise and the integration of information from the different EEG sensors. The proposed algorithm provides a natural solution for both problems. By combining the different sensors using alternating diffusion we obtain non-linear filtering abilities, reveal the common structures, which represent underlying properties of the EEG data, and significantly attenuate (sensor-specific) noise and interferences.

4. REPRESENTING EEG DATA USING ALTERNATING DIFFUSION

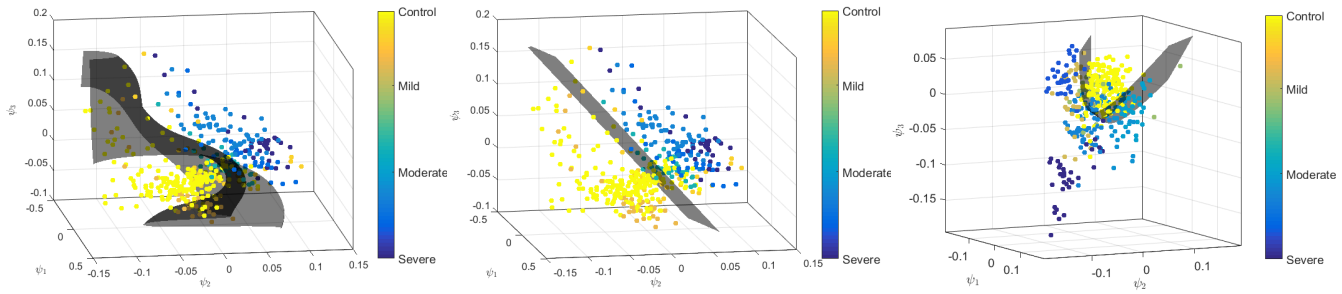
We applied alternating diffusion to EEG recordings of Alzheimer's disease patients and healthy subjects in order to assess disease severity. We analyzed EEG recordings of 26 subjects in resting state. Two different EEG devices were used to record the EEG signals, each containing 64 electrodes. Due to significant differences in noise levels between the output of the two devices, recordings from each device were analyzed separately. Details regarding the two subject groups, recorded by the different devices, appear in Table 1. The disease state of the subjects was determined based on the Mini Mental State Examination (MMSE) test. The MMSE score associated with each disease state appears in Table 1.

An initial pre-processing stage was performed in which a band-pass filter (BPF) and a notch filter at 50[Hz] were applied. The pass-band frequencies used for each EEG device are presented in Table 2. The BPF frequencies were selected such that frequencies related to alpha, beta, gamma and delta bands are mostly preserved and frquencies with dominant noise are filtered. These frequency bands have been shown to contain significant differences in Alzheimer patients compared with healthy subjects [21].

After the pre-processing stage, recordings from each subject were divided into short time segments (see Table 2, for exact time frames/window sizes), and the Scattering Transform [22] was applied to each time segment, in each electrode separately. The Scattering Transform is a translation invariant transform which was shown to be stable to time deformations; it is computed by applying a cascade of wavelet and modulus operators. This transform builds an invariant and informative signal representation and was applied in order to obtain meaningful features. For each EEG electrode and each EEG device, the Scattering Transforms of all subjects were concatenated to obtain a matrix of $m \times N$ elements, where m is the dimension of the Scattering Transform output and N is the number of time segments from all subjects recorded by the same EEG device. Note that we are combining recordings of different subjects from the same electrode, although the exact position of the electrode might slightly change due to movement of the EEG cap. In this setting we neglect these differences and assume that the data from different subjects can be combined. We further note that the scatter plots presented in Figure 1 and Figure 2 depict that our framework indeed overcomes these differences and embed subjects based on their disease state.

Based on the resulting $m \times N$ matrix, the affinity kernel described in (1) was constructed for each electrode. This resulted in kernel matrices of size $N \times N$, in which each matrix entry represents the affinity between two time segments, based on data from one EEG electrode.

Finally, alternating diffusion was applied to recordings from each device separately. For each device, several combinations of 2 or 3 electrodes were examined by multiplying the normalized kernels (Markov matrices) (2) corresponding to the chosen electrodes and then by calculating the eigenvectors of the resulting matrix. Figure 1 and Figure 2 present exemplary results of the two EEG devices. The labels of the used electrodes (channels) from each device are presented in Table 2. These electrodes were chosen empirically, however, several different combinations led to similar satisfactory



(a) (b) (c)

Fig. 3: Examples of different SVM hyperplanes which separate the control group and Alzheimer group. (a) Application of SVM with a radial basis function to the embedding of the first EEG device. (b) Application of linear SVM to the embedding of the first EEG device. (c) Application of SVM with a quadric function to the embedding of the second EEG device.

Table 3: Comparison of the classification based on alternating diffusion coordinates with diffusion maps and the scattering transform.

	Alternating Diffusion	Diffusion Maps	Scattering Transform
Healthy Subjects	**10/11**	7/11	5/11
Alzheimer Patients	**13/15**	11/15	10/15

results. We note that only a small number of electrodes was considered since alternating diffusion reveals only common features in all sensors. Therefore, when combining too many electrodes, informative features might be lost.

In Figure 1 and Figure 2, 3D scatter plots of the first three eigenvectors calculated by alternating diffusion are presented. Each point in these plots corresponds to one time segment of a subject and is colored according to the MMSE score of that subject. Figure 1 presents the data from Device 1 and Figure 2 presents the data from Device 2. Both figures depict that the new coordinate system, obtained by alternating diffusion, creates an embedding, which separates different disease states to different locations in the constructed 3D space. We note that in this embedding, healthy subjects are grouped together and Alzheimer patients are more dispersed. This is expected due to large variability between different Alzheimer patients, even when diagnosed with the same MMSE score. We emphasize that the plots in Figure 1 and Figure 2 were constructed in a purely data-driven manner, without prior knowledge on the true subject classification. Therefore, these figures illustrate that alternating diffusion can be applied to EEG recordings in order to obtain a new coordinate system, in which Alzheimer patients and healthy subjects can be classified based on Euclidean proximity. Furthermore, based on the embedded coordinates, a classification to different disease states can be obtained, as most subjects belonging to the same class are grouped together in the 3D space.

In order to assess the new embedding and visualize its advantages, for each EEG device separately, we applied a support vector machine (SVM) classifier to the first 3 eigenvectors, corresponding to the 3 largest eigenvalues, calculated by alternating diffusion. We visualize the resulting hyperplanes in Figure 3. Figure 3 presents 3 examples of different SVM hyperplanes separating the control group and Alzheimer group in the new coordinate system. The first two plots present the application of SVM with a radial basis function and linear SVM to the embedding of the EEG recordings from the first device and the third plot presents the application of SVM with a quadric function to the embedding obtained from the second EEG device. These plots depict that by applying such standard classifiers to the constructed embedding, a good separation between the control group and Alzheimer group can be obtained in both devices.

Leave-one-out cross validation was performed using an SVM classifier with a radial basis function (RBF), for each EEG device. At each iteration, all the time-segment points of one subject were removed and used as a test-set. The class of each test subject was then determined based on the majority class of the time-segment points. This resulted in a correct classification of 10 out of the 11 control subjects and 13 out of the 15 Alzheimer patients. Moreover, when applying multiclass SVM, classification of the different disease stages can also be achieved based on the constructed coordinates, in a purely data-driven manner.

In order to establish the advantages of the new embedding, we compared the classification based on the alternating diffusion coordinates of multiple EEG channels, with classifications based on diffusion maps of multiple EEG channels (concatenated) and a classification based on the scattering transform output (without further analysis). In all instances, alternating diffusion outperformed all other methods in the classification of the disease states and control group, as presented in Table 3.

5. CONCLUSION

Alternating diffusion is a data-driven method for extracting the common source of variability from multiple sensor measurements. We demonstrated the ability of alternating diffusion to extract the common variable from EEG recordings of Alzheimer patients and healthy subjects, affected by measurement noise, nuisance movement and non-linearities due to the cranium. We show that alternating diffusion provides a new coordinate system in which the subjects are divided into groups of healthy subjects and Alzheimer patients without prior knowledge on the disease.

In future work, we plan to extend this study and include more subjects in order to validate our results on a greater scale. In addition, we plan to devise a new framework which considers neuronal networks by using a larger number of EEG electrodes.

6. REFERENCES

[1] S. Barrack, *Biochemical Biomarkers in Alzheimer's Disease*, iMedPub, 2013.

[2] W. Henry, F. L. Querfurth, and F. M. LaFerla, "Alzheimer's disease," *N Engl J Med*, vol. 362, pp. 329–44, 2010.

[3] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of alzheimer's disease report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer's disease," *Neurology*, vol. 34, no. 7, pp. 939–939, 1984.

[4] J. Jeong, "Eeg dynamics in patients with alzheimer's disease," *Clinical neurophysiology*, vol. 115, no. 7, pp. 1490–1505, 2004.

[5] M. Ahmadlou, H. Adeli, and A. Adeli, "New diagnostic eeg markers of the alzheimer's disease using visibility graph," *Journal of neural transmission*, vol. 117, no. 9, pp. 1099–1109, 2010.

[6] C. Lehmann, T. Koenig, V. Jelic, L. Prichep, R. E. John, L.O. Wahlund, Y. Dodge, and T. Dierks, "Application and comparison of classification algorithms for recognition of alzheimer's disease in electrical brain activity (eeg)," *Journal of neuroscience methods*, vol. 161, no. 2, pp. 342–350, 2007.

[7] A. Cichocki, S. L. Shishkin, T. Musha, Z. Leonowicz, T. Asada, and T. Kurachi, "Eeg filtering based on blind source separation (bss) for early detection of alzheimer's disease," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 729–737, 2005.

[8] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A spatio-temporal wavelet-chaos methodology for eeg-based diagnosis of alzheimer's disease," *Neuroscience letters*, vol. 444, no. 2, pp. 190–194, 2008.

[9] B. N. Cuffin, "Effects of local variations in skull and scalp thickness on eeg's and meg's," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 1, pp. 42–48, 1993.

[10] R. R. Lederman and R. Talmon, "Learning the geometry of common latent variables using alternating-diffusion," *Applied and Computational Harmonic Analysis*, 2015.

[11] R. Talmon and H. T. Wu, "Latent common manifold learning with alternating diffusion: analysis and applications," *arXiv preprint arXiv:1602.00078*, 2016.

[12] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, pp. 5–30, Jul. 2006.

[13] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps, spectral clustering and eigenfunctions of fokker-planck operators," *Neural Information Process. Systems (NIPS)*, vol. 18.

[14] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps, spectral clustering and reaction coordinates of dynamical systems," *Appl. Comput. Harmon. Anal.*, pp. 113–127, 2006.

[15] R. R. Lederman, R. Talmon, H. T. Wu, Y. L. Lo, and R. R. Coifman, "Alternating diffusion for common manifold learning with application to sleep stage assessment," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 5758–5762.

[16] R. Talmon, S. Mallat, H. Zaveri, and R. R. Coifman, "Manifold learning for latent variable inference in dynamical systems," *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3843–3856, 2015.

[17] V. Chudacek, R. Talmon, J. Anden, S. Mallat, R.R. Coifman, P. Abry, and M. Doret, "Low dimensional manifold embedding for scattering coefficients of intrapartum fetale heart rate variability," in *Proc. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 6373–6376.

[18] Y. Zhang, G. Xu, J. Wang, and L. Liang, "An automatic patient-specific seizure onset detection method in intracranial eeg based on incremental nonlinear dimensionalityreduction," *Computers in biology and medicine*, vol. 40, no. 11, pp. 889–899, 2010.

[19] J. Kortelainen, E. Vayrynen, and T. Seppanen, "Isomap approach to eeg-based assessment of neurophysiological changes during anesthesia," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 2, pp. 113–120, 2011.

[20] Y. Pan, S. S. Ge, A. Al Mamun, and F. R. Tang, "Detection of seizures in eeg signal using weighted locally linear embedding and svm classifier," in *Proc. IEEE Conference on Cybernetics and Intelligent Systems*, 2008, pp. 358–363.

[21] M. Penttilä, J. V. Partanen, H. Soininen, and P. J. Riekkinen, "Quantitative analysis of occipital eeg in different stages of alzheimer's disease," *Electroencephalography and clinical neurophysiology*, vol. 60, no. 1, pp. 1–6, 1985.

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