Earthquake-explosion discrimination using diffusion maps
Neta Rabin,¹ Yuri Bregman,² Ofir Lindenbaum,³⁴ Yochai Ben-Horin,² and
Amir Averbuch,³

¹Department of Basic Sciences, Afeka -
Tel Aviv Academic College of Engineering,
Tel-Aviv, Israel.

²Soreq Nuclear Research Center, Yavne,
81800, Israel.

³Department of Computer Science,
Tel-Aviv University, Israel.

⁴Department of Electrical Engineering,
Tel-Aviv University, Israel.
Key Points.

Seismic event discrimination and test-ban treaty verification, Dimensionality reduction, Diffusion Maps, Time series analysis

Abstract.

Discrimination between earthquakes and explosions is an essential component of nuclear test monitoring and it is also important for maintaining the quality of earthquake catalogs. Currently used discrimination methods provide a partial solution to the problem. In this work, we apply advanced machine learning methods and in particular diffusion maps for modeling and discrimination of seismic signals. Diffusion maps enable us to construct a geometric representation that capture the intrinsic structure of the seismograms. The diffusion maps are applied after a pre-processing step, in which seismograms are converted to normalized sonograms. The constructed low-dimensional model is used for automatic earthquake-explosion discrimination of data that is collected in single seismic stations. We demonstrate our approach on a data set comprising seismic events from the Dead Sea area. The diffusion-based algorithm provides correct discrimination rate that is higher than 90%.
1. Introduction

Discrimination between earthquakes and explosions is an essential component of nuclear test monitoring and it is also important for maintaining high-quality earthquake catalogs. Typical discrimination algorithms calculate several seismic parameters from the input seismograms and use these parameters to distinguish between earthquakes and explosions. The simplest seismic parameter is focal depth. Its drawback is that estimation of focal depth is usually inaccurate in lack of depth phases. Other widely used seismic discrimination methods are Ms:mb (surface wave magnitude versus body wave magnitude) and spectral amplitude ratios of different seismic phases [Blandford, 1982; Rodgers, 1997]. However, discrimination methods based on seismic parameters give only a partial solution to the problem. For instance, 54987 (60%) out of 90967 seismic events were reported by the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) in years 2011-2013 without conclusive identification as earthquakes while it is clear that most of those 54987 events are typically earthquakes [Ben Horin, 2015]. Regional earthquake catalogs are often contaminated with explosions which may cause the erroneous estimation of a seismicity hazard [Kortstrom, 2016].

Over the last two decades different machine learning algorithms were applied to classify seismic events in an unsupervised or in a supervised mode. These include the application of artificial neural networks [Tiira, 1996; Del Pezzo, 2003; Esposito, 2006], self-organizing maps [Kohler, 2010; Kuyuk, 2011; Sick, 2015], hidden Markov models [Ohrnberger, 2001; Beyreuther, 2008; Hammer, 2013] and support vector machines [Kortstrom, 2016].
In this paper we propose to apply a graph-based machine learning tool and in particular the diffusion maps method [Coifman, 2006] for organizing a large number of events that are captured in a seismic station and for classifying new recorded events. Diffusion maps were applied in seismology for the estimation of arrival times [Taylor, 2011] and for the seismic phase classification [Ramirez, 2011], but for the best of our knowledge the method has not been used for event discrimination.

The proposed algorithm begins with a preprocessing stage in which a time-frequency representation is extracted from each seismic event. The training phase includes the construction of a normalized graph that holds the local connections between the seismic events. A low dimensional map is then obtained by the eigendecomposition of the graph. The constructed embedding is distance preserving, thus the geometry of the dataset is kept in the new embedding coordinates. The constructed low-dimensional representation can be extended to include new gathered data points by application of geometric harmonics [Coifman, 2006]. Once a new seismic event is included in the low-dimensional representation, it is classified according to the known labels of the training data. Here, the k-Nearest Neighbors (k-NN) algorithm is applied for classification of new events.

We demonstrate our approach on a data set comprising seismic events from the Dead Sea area that were taken from the seismic catalog of the Geophysical Institute of Israel for years 2004-2014 with duration magnitudes $\text{Md} \geq 2.5$. We used the waveforms of two seismic stations on local distances. The diffusion-based algorithm provides correct discrimination rate that is higher than 90%.

The paper is organized as follows: Section 2 reviews the mathematical methods that are used for dimensionality reduction. In Section 3 we describe how the proposed framework
is applied discriminate between earthquakes and explosions for the single seismic station.

Discrimination results are described in Section 4. We conclude and propose possible extensions of this work in Section 5.

2. Mathematical Methods

Advanced machine learning methods, which aim to model and learn large and complex datasets, often include the construction of a graph to describe the local relationship between the data points. A compact representation of this graph is obtained by a spectral decomposition, where the eigenvectors of the graph embed the data into a low dimensional space. Such methods are known as dimensionality reduction or manifolds learning techniques. Local linear embedding [Roweis, 2000], Laplacian eigenmaps [Belkin, 2003, 2004] and diffusion maps [Coifman, 2006] are a common manifold learning techniques. In this work, the diffusion maps framework is applied for embedding the original high-dimensional data to a low dimensional manifold while preserving intrinsic geometry of the observed data set.

2.1. Diffusion Maps

Let \( X = \{x_1, \ldots, x_M\} \) be a set of points, where \( x_i \in \mathbb{R}^D \). A graph that includes the data points \( X \) as its nodes and a kernel \( W \triangleq w(x_i, x_j) \) for its weighted edges is constructed. The kernel \( W \) measures the pairwise similarity between the data points and satisfies the following properties:

- \( W \) is symmetric: \( w(x_i, x_j) = w(x_j, x_i) \);
- \( W \) is positive-preserving: \( w(x_i, x_j) \geq 0 \) for all \( x_i \in X \);
\( W \) is positive semi-definite: for all real-valued bounded function \( f \) defined on \( X \), it satisfies \( \sum_i \sum_j w(x_i, x_j) f(x_i) f(x_j) \geq 0 \).

A typical choice for the kernel is the Gaussian kernel \( W = w(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\epsilon} \). The kernel is normalized to be a Markov matrix. The normalization is also known as the graph Laplacian (see Chung [1997]). It is computed by

\[
P \triangleq p(x_i, x_j) = s(x_i)^{-1} w(x_i, x_j),
\]

where \( s(x_i) = \sum_{x_j \in X} w(x_i, x_j) \).

The Markov transition matrix \( P \) holds the probabilities of moving from \( x_i \) to \( x_j \) in one time step. The eigendecomposition of \( P \) is computed by

\[
p(x_i, x_j) = \sum_{k \geq 0} \lambda_k \psi_k(x_i) \phi_k(x_j),
\]

where \( \{\lambda_k\}_{k=0}^{M-1} \) are the eigenvalues of \( P \) and \( \{\psi_k\}_{k=0}^{M-1}, \{\phi_k\}_{k=0}^{M-1} \) are the corresponding left and right biorthogonal eigenvectors of \( P \). From the properties of the matrix \( P \) (see Chung [1997]), the spectrum \( \{\lambda_k\} \) decays fast to zero, thus, only a few terms \( d << D \) are required to achieve sufficient accuracy in Equation (2). These few leading modes are used to embed the data into a low dimensional space. The family of diffusion maps \( \{\Psi(x_i)\}_{i=1}^{M-1} \) are defined by

\[
\Psi(x_i) = (\lambda_1 \psi_1(x_i), \lambda_2 \psi_2(x_i), \lambda_3 \psi_3(x_i), \cdots).
\]

Note that the first eigenvector \( \psi_0 \) is not included in the diffusion maps definition, as it is constant. Next, we review the diffusion distance [Coifman, 2006] and show that diffusion maps embed the dataset into a Euclidean space and pairwise distances between the datapoints in the original space are preserved in the embedding process.
The **diffusion distance** between two data points \( x_i \) and \( x_j \) is defined as the weighted \( L^2 \) distance

\[
D(x_i, x_j)^2 = \sum_{x_m \in X} \frac{(p(x_i, x_m) - p(x_m, x_j))^2}{\phi_0(x_m)},
\]

(4)

In this metric, two data points are close to each other if many existing paths connect them. The value of \( \frac{1}{\phi_0(x_m)} \) depends on the point’s density. Using the spectral decomposition of \( P \) in Equation (2) to describe the nominator of Equation (4) together with the biorthogonality of the left and right eigenvectors of \( P \), the diffusion distance is expressed by

\[
D(x_i, x_j)^2 = \sum_{k \geq 1} \lambda_k^2 (\psi_k(x_i) - \psi_k(x_j))^2,
\]

(5)

where \( \{\psi_k\} \) are the right eigenvectors of \( P \). Thus, the Euclidean distance between two points in the embedded space is equivalent to the distances between the points as defined by a random walk and the embedding preserves the geometry (pairwise distances) of the data points.

### 2.1.1. Setting the Scale Parameter \( \epsilon \).

When constructing the Gaussian kernel \( W \) in Equation (1), one has to set the value of the scale (width) parameter \( \epsilon \). Setting \( \epsilon \) to be too small may result in very small local neighborhoods that are not able to capture the local structure around the point. On the contrary, setting \( \epsilon \) to be too large may result in a fully connected graph that may generate a coarse description of the data.

We seek for a value of \( \epsilon \) that would be most effective for our classification task. First, we describe a scheme proposed by Singer et al. [Singer, 2009]. Their scheme aims to find a range of values for \( \epsilon \). The idea is to compute the kernel at various values of \( \epsilon \) and...
search for the range of values where the Gaussian bell shape exists. The proposed scheme is outlined in algorithm 1.

**Algorithm 1 $\epsilon$ range selection**

**Input:** A dataset $X = \{x_1, x_2, \ldots, x_M\}, x_i \in \mathbb{R}^D$.

**Output:** A range of values for $\epsilon$, $\hat{\epsilon} = [\epsilon_0, \epsilon_1]$.

1. Compute Gaussian kernels $W(\epsilon)$ for several values of $\epsilon$.
2. Compute: $L(\epsilon) = \sum_i \sum_j W_{ij}(\epsilon)$ for these values.
3. Plot a logarithmic plot of $L(\epsilon)$ (vs. $\epsilon$).
4. Set $\bar{\epsilon}$ as the maximal linear range of $L(\epsilon)$.

Note that the two asymptotes would always be $L(\epsilon) \xrightarrow{\epsilon \to 0} \log(M)$, and $L(\epsilon) \xrightarrow{\epsilon \to \infty} \log(M^2) = 2\log(M)$, since for $\epsilon \to 0$, $W$ approaches the Identity matrix, and for $\epsilon \to \infty$, $W$ approaches an all-ones matrix.

In this study we apply a scheme proposed by [Lindenbaum, 2015] for a supervised selection of $\epsilon$ dedicated for classification tasks. Given a training set $X$ and given two known classes $C_1, C_2$, with $G_1, G_2$ data points within each class, compute several diffusion maps for various values of $\epsilon$ within the range that is computed using Algorithm 1. Using the desired amount of leading coordinates $d << D$ within each computed diffusion map $\Psi(\epsilon)$, denote $\mu_1, \mu_2$ as the center of mass of classes, computed in the diffusion coordinates, and $\mu_a$ as the center of mass of all data points. Compute the average square distance of the $M$ data points from the center of mass within the class:

$$D_{\epsilon_i} = \frac{1}{G_i} \sum_{x_m \in C_i} ||\Psi(x_m) - \mu_i||^2, \ i = 1, 2.$$  (6)
Then, compute the same measure for all of the data:

\[
D_a = \frac{1}{M} \sum_{x_m \in X} ||\Psi(x_m) - \mu_a||^2.
\]  

(7)

Finally, find \( \epsilon \) that minimizes the ratio:

\[
\hat{\epsilon} = \arg \min_\epsilon \sum_{i=1}^{2} \frac{D_{c_i}}{D_a}.
\]  

(8)

This \( \epsilon \) inherits the inner structure of the classes and neglects the mutual structure.

2.2. Geometric Harmonics

The construction of diffusion map coordinates requires a spectral decomposition of an \( M \times M \) matrix, where \( M \) is the number of data points. When \( M \) is large (includes for example tens of thousands of data points) this procedure is time consuming and therefore we do not want to apply it for each new arriving sample. Even though the dataset that is used in this paper is small, we outline an extension procedure that can be deployed for online classification of new events. The method is applied in Section 4.

Extension of non-linear embedding coordinates is traditionally done by using the Nyström method [Nyström, 1929; Miller, 1974]. Other multi-scale methods for extension include spectral multi-scale function extensions [Bermanis, 2013] or Laplacian Pyramids [Rabin, 2012]. In this paper, the geometric harmonics method [Coifman, 2006] that is based on the Nyström technique is applied. The method involves construction of harmonic functions that form a basis and thus can span a given function \( f \). In particular, the harmonic functions are used to span and extend the diffusion maps embedding coordinates.
Let $X$ be a set of points in $\mathbb{R}^D$ and $f(x) = \lambda_1 \psi_1(x)$ be the corresponding first diffusion maps embedding map. Let $\bar{x} \in \mathbb{R}^D$ be a new data point. The geometric harmonics scheme extends $f(x) = \lambda_1 \psi_1(x)$ to $\bar{x}$, it evaluates $f(\bar{x}) = \lambda_1 \psi_1(\bar{x})$.

We form a basis that can span any given function $f$. The basis is composed of eigenvectors associated with a kernel matrix. A kernel (typically a Gaussian kernel) with a bandwidth scale equal to $\sigma > 0$ is constructed from the dataset $X$. The parameter $\sigma$ is the scale of the extension, which depends on the smoothness of the approximated function (see [Coifman, 2006]). In general it is desirable that $\sigma$ would be larger than $\epsilon$, which was the kernel width that was set in diffusion maps. Denote this kernel by $W_\sigma = w_\sigma(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma}}$ and compute its spectral decomposition. The eigenvalues and eigenvectors of $W_\sigma$ are denoted by $\{\mu_k\}_{k=0}^{M-1}$ and $\{\varphi_k\}_{k=0}^{M-1}$, they satisfy

$$\mu_k \varphi_k(x_i) = \sum_{x_j \in X} e^{-\frac{\|x_i - x_j\|^2}{\sigma}} \varphi_k(x_j), \ x_i \in X.$$ (9)

Now, a general function $f$ and in particular $f(x) = \lambda_1 \psi_1(x)$, can by approximated as a linear combination of this basis by

$$\lambda_1 \psi_1(x_i) = f(x_i) = \sum_{k: \mu_k \geq \eta \mu_0} \langle \varphi_k, f \rangle \varphi_k(x_i), \ x_i \in X$$ (10)

Since the eigenvalues $\{\mu_k\}$ tend to zero as $k \to \infty$, the scheme (9) is ill-conditioned. In order to avoid this ill-conditioning, the geometric harmonics cuts off the sum in Equation (10) by introducing a parameter $\eta$ and only using the eigenvectors that satisfy $\mu_k \geq \eta \mu_0$.

Note that this procedure, which includes construction of the kernel $W_\sigma$ and calculation of its eigenvalues and eigenvectors, is only carried out once as a pre-processing step. Evaluating the embedding location of a new point $\bar{x}$ in an online mode only requires the two following steps (Equations 11 and 12), which are fast to compute. The functions in...
the basis \( \{ \varphi_k \}_{k: \mu_k \geq \eta \mu_0} \) are extended to the point \( \bar{x} \) by

\[
\tilde{\varphi}_k(\bar{x}) = \frac{1}{\mu_k} \sum_{x_j \in X} e^{-\frac{||\bar{x} - x_j||^2}{\sigma}} \varphi_k(x_j).
\] (11)

The diffusion coordinate \( f = \lambda_1 \psi_1 \) is extended to the new point \( \bar{x} \) by

\[
\lambda_1 \psi_1(\bar{x}) = \tilde{f}(\bar{x}) = \sum_{k: \mu_k \geq \eta \mu_0} \langle \varphi_k, f \rangle \tilde{\varphi}_k(\bar{x}).
\] (12)

Extension of the next diffusion maps coordinates \( \lambda_2 \psi_2, \lambda_3 \psi_3, \ldots \) is done in a similar manner, with the use of the same basis \( \{ \mu_k \}, \{ \varphi_k \} \).

### 3. Application of Diffusion Maps for Seismic Discrimination

In this section we describe how the proposed diffusion maps framework is applied to the problem of seismic discrimination. First, the dataset is specified. Next, the pre-processing phase that results in a time-frequency representation for each seismic waveform is described. Finally, we outline the steps of a training and testing algorithm for learning a given data set and discriminating new collected events.

#### 3.1. Description of the Data Set

A dataset that includes 44 earthquakes and 62 explosions that were recorded at two different stations was constructed for the feasibility study. The collected waveforms occurred in the Dead Sea area between the years 2004-2014. Figure 1 presents a map with the locations of the collected data. The exact locations are between latitudes 31\(^\circ\)N-32\(^\circ\)N and longitudes 34.9\(^\circ\)E-35.7\(^\circ\)E. The duration magnitudes of all events are \( \text{Md} \geq 2.5 \). The waveforms were independently analyzed by the Geophysical Institute of Israel (GII), thus, we have a reliable label (earthquake or explosion) for each event.

Most of earthquakes in the dataset occurred in the Dead Sea basin, which is the largest fault of the Dead Sea [Garfunkel, 2014]. The earthquakes dataset includes in it an event
from February 11, 2004 that is of duration magnitude of Md = 5.1. This was the strongest event in this area since 1927 [Hofstetter, 2008]. Twelve aftershocks that are included in the dataset are associated with this main shock. Their magnitudes are between 2.5-3.7 magnitude units.

Many of the explosions in the dataset originate in the cluster of Negev phosphate queries that is located in the south-eastern part of the area (see Figure 1). These explosions are ripple-fire query blasts. The dataset also includes two experimental underwater explosions in the Dead Sea that were conducted by the GII. Their duration magnitudes are 3.0 and 3.1 [Hofstetter, 2008].

The average magnitude durations of the earthquakes and the explosions is 3.2 and 2.7 respectively. A histograms plot of the magnitude distribution is presented in Figure 2. The focal depths of earthquakes is between 4 and 23 km with a mean of 13 km.

The dataset consists of recordings from two different broad band stations MMLI (Malkishua) and HRFI (Harif). Both stations are operated by the GII and they are part of the Israel National Seismic Network [Hofstetter, 2008]. MMLI and HRFI are dual use stations, they are also part of the Israel Cooperating National Facility (CNF) as defined by the Comprehensive Nuclear-Test Ban Treaty (CTBT) [Bartal, 2000]. MMLI and HRFI stations are located to the north and to the south of the analyzed region, respectively (see Figure 1). The mean distances from MMLI and HRFI to the earthquakes are 111 km and 116 km, respectively, and the mean distances to explosions are 145 km and 125 km, respectively. Each station is equipped with a three component STS-2 seismometer and a Quanterra data logger. The sampling frequency is 40 Hz.
Each waveform in the dataset is a six minutes long recording that consist of 10,000 samples. The first P phase onset resides 10 seconds after the beginning of each waveform.

3.2. Feature Extraction by Normalized Sonograms

The feature extraction step constructs a time-frequency representation from the given waveforms. The proposed method is based on sonograms (Joswig, 1990) with some modifications. It enables us to characterize a transient seismic signal by its time varying spectra. Each single-trace seismic waveform, denoted by $y[n] \in \mathbb{R}^N$ is a time series signal of length $N = 10,000$ sampled at the rate of $F_s = 40Hz$. The waveform $y[n]$ is decomposed into a set of overlapping windows of length $N_0 = 256$. In this paper the overlap ratio is set to $s = 0.8$, which yields that the number of time windows is equal to $T = 192$. The short-time Fourier transform (STFT) is applied to $y[n]$ in each time window, which is preliminary Hanning-tapered. Then, power spectral densities are computed. The resulting spectrogram is denoted by $R(f,t)$, where $f$ is a frequency bin and $t$ is a number of time window (bin). Thus, it contains $T = 192$ time bins and $N_0/2 = 128$ frequency bins.

The sonogram is obtained by summing the spectrogram in logarithmically scaled frequency pass bands for every time bin. The frequency scale is rearranged to be equally tempered on a logarithmic scale into near half-octave logarithmically scaled frequency pass bands. The following 10 frequency pass bands are used in this study: 0.157-0.315 Hz, 0.315-0.630 Hz, 0.630-1.102 Hz, 1.102-1.889 Hz, 1.889-2.992 Hz, 2.992-4.567 Hz, 4.567-6.772 Hz, 6.772-9.921 Hz, 9.921-14.331 Hz and 14.331-20 Hz. Finally the sonogram is normalized such that the sum of energy in every frequency band is equal to 1. The result is a normalized sonogram and it is denoted by $S(k,t)$, where $k$ is the frequency
band number and $t$ is the time window number. Note that a similar normalization of sonograms was used in [Ohrnberger, 2001].

Figure 3 illustrates the feature extraction process. The top part displays four seismograms that were recorded at the vertical channel of station MMLI. The first three (from top) seismograms belong to earthquakes in the northern Dead Sea area with duration magnitudes $Md$ of 5.1, 2.6 and 3.5, respectively, the bottom seismogram belongs to an underwater explosion with duration magnitude $Md = 3$. The maximal distance between the epicenters of four events is 6.5 km. The bottom part presents the resulting normalized sonograms. Notice that the original seismograms have completely different amplitude ranges. For instance, the amplitude ratio between the first and the second seismograms is about $10^6$. However the normalized sonograms transform all events in similar ranges of amplitudes.

3.3. The training step

In this section we describe the steps required for constructing a low dimensional space which inherits the desired properties for the discrimination task. In this study a training set that includes approximately hundred events is used, however the approach could be implemented for much larger dataset. Algorithm 2 describes the proposed training procedure.
**Algorithm 2** The training algorithm

**Input:** A dataset of seismic waveforms \( y_1, ..., y_M \).

**Output:** A low dimensional mapping that embeds the waveforms into a Euclidean space \( \Psi^Y \).

1. For each waveform \( y_i \), compute the spectrogram \( R_i(f, t), i = 1, ..., M \).
2. Use \( R_i(f, t), i = 1, ..., M \) to compute the normalized sonograms \( S_i(k, t), i = 1, ..., M \).
3. Reshape \( S_i(k, t) \) into \( x_i \) by concatenating columns.
4. Compute optimal scale \( \hat{\epsilon} \).
5. Compute the diffusion mapping \( \Psi(x_i), i = 1, ..., M \).
6. Set the \( d << M \) leading coordinates of \( \Psi(x_i), i = 1, ..., M \) to embed the data.

For the last three steps, we apply Algorithm 1 to compute a range of valid values for \( \epsilon \), then compute the optimal value based on Eqs. (6), (7) and (8). This results in an optimal kernel scale for Step 4. Next, we use the dataset \( X = \{ x_1, x_2, ..., x_M \} \) to compute a Gaussian kernel \( W \). It is normalized into a transition matrix \( P \), as explained in Eq. (1).

Computing the spectral decomposition of \( P \) results in the set of eigenvalues \( \{ \lambda_k \}_{k=0}^{M-1} \) and eigenvectors \( \{ \psi_k \}_{k=0}^{M-1} \). This spectral decomposition yields the set of diffusion map coordinates for Step 5, as given in Eq. (3).

Last, \( \Psi(S_i) \) is truncated in Step 6 by choosing \( d \) leading coordinates. The value of \( d \) is chosen by examining the eigenvalues of \( P \), a typical choice is such that \( \lambda_{d+1} < 0.05 \).

### 3.4. The classification step

The low dimensional representation that is constructed in the Training step is used for classifying new data points with an unknown category. The new data point is embedded
in the low dimensional space using Geometric harmonics, and classification is performed using k-NN. The steps are described in the following algorithm.

**Algorithm 3** The classification algorithm

**Input:** A low dimensional representation of the train data $\Psi(X)$, a vector of labels $z \in \mathbb{R}^M$, a test waveform $\hat{y}$

**Output:** A label $\hat{z}$ for the query point.

1. Compute the normalized sonogram $\hat{S}(k,t)$ from the seismic waveform $\hat{y}$.
2. Reshape $\hat{S}(k,t)$ into $\hat{x}$ by concatenating columns.
3. Extend the diffusion coordinates $\Psi(\hat{x})$ to include $\hat{x}$.
4. Classify the point $\Psi(\hat{x})$ in the diffusion space and determine the label $\hat{z}$.

Steps 1-2 from Algorithm 2 to compute the feature vector $\hat{S}$ for every test point $\hat{y}$.

The Geometric Harmonics method that is described in Section 2.2 is to compute the approximated diffusion coordinates $\Psi(\hat{x})$ in Step 3. In Step 4 we find the k-Nearest Neighbors of $\Psi(\hat{x})$ in the diffusion maps space and use their class label to determine whether $\hat{y}$ is a natural or an unnatural event.

4. Experimental Results

4.1. A first Example

This first example demonstrates the application of diffusion maps on a subset of the dataset. The subset comprises of 20 events that occurred in the northern Dead Sea area within latitudes $31.6^\circ$N- $32^\circ$N and longitudes $35.3^\circ$E-$35.6^\circ$E. The first 19 events are earthquakes with duration magnitudes ranging between 2.5-5.1. The last event is an explosion of magnitude 3. Each event is pre-processed and replaced by a normalized sonogram (see Steps 1-2 of Algorithm 2). Next, the diffusion maps algorithm is applied as
explained in Algorithm 2 Steps 3-5. A kernel matrix $W$ of size $20 \times 20$ is constructed and normalized. The geometric harmonics (Section 2.2) is not applied in this first example, as we mainly wish to demonstrate the method’s ability to organize the high dimensional data in a low dimensional space according to its intrinsic physical properties.

Figure 4 (left) shows the low-dimensional embedding of the dataset $Y$ into the first two diffusion maps coordinates. We clearly see that the first embedding coordinate (x-axis) separates the explosion from 19 earthquakes. Furthermore, in Figure 4 (right), the events are colored by their magnitude. The second diffusion maps coordinate (y-axis) highly correlated with the magnitude, as the color changes quite smoothly. Recall that these meta-parameters: longitude, latitude, depth and event-type are not part of the algorithm’s input. They obviously effect the geometric organization of the events (sonograms) in the original high dimensional space, and this natural organization is preserved by diffusion maps.

4.2. Discrimination Results

This section describes discrimination experiments performed using our framework. We use the full dataset that contains 106 events, which were recorded at two 3-component seismic stations. Thus, we have six datasets, each of size 106.

4.3. Low Dimensional Mapping of Seismic Events

Given a dataset that belongs to one channel of one station, $Y = \{y_i\}_{i=1}^{106}$, each event $y_i$ is pre-processed and replaced by a normalized sonogram denoted by $S_i(f, t) \in \mathbb{R}^{192 \times 10}$ (Sonograms were constructed using 192 time windows and 10 frequency bands). Each normalized sonogram, is transformed into a one column vector $x_i \in \mathbb{R}^{1920}$ by concatenating
the columns (this process is explained in Algorithm 2). The diffusion maps is applied to the set of sonograms. The optimal scale parameter $\hat{\epsilon}$ (Equation (8)) is computed using the method described in 2.1.1. Then, we compute a low-dimensional mapping for each sensor based on Eq. (1). The mappings are denoted by $\Psi^{HZ}$, $\Psi^{HE}$, $\Psi^{HN}$, $\Psi^{MZ}$, $\Psi^{ME}$, $\Psi^{HN}$ for vertical, east and north components of station HRFI and MMLI, respectively. The two dimensional mappings of the data from HRFI are presented in Figure 5. The separation between the two classes is clearly evident in the diffusion maps coordinates.

4.4. Event Classification

In order to evaluate how effective the diffusion mapping has separated the earthquakes from the explosions we use a standard leave-one-out cross validation technique. The diffusion maps is computed 106 times, leaving 1 test point out at each iteration. The embedding coordinates of the test points are then evaluated with the geometric harmonics algorithm (see Step 2 of Algorithm 3). Label classification for the test point is performed using a k-Nearest Neighbors classifier. We set $k = 1$. Using the 1-NN the estimated class is simply the class of the nearest data point (Euclidean distance). In Figure 6 a 3-dimensional diffusion mapping based on sensor Z of HRFI is presented, the color represents the classification results (only 3 explosions are misclassified as earthquakes in this experiment).

The performance of the diffusion maps classifier is compared with the same type of classifier (1-Nearest Neighbor) that is based on Principal Component Analysis [Wold, 1987]. The main difference between the algorithms is the way the dimension of the sonogram dataset is reduced. PCA is a linear dimensionality reduction method that projects the data into new coordinates, the new coordinates are linear combinations of the original
coordinated. In diffusion maps the data is embedded, rather than projected, into a new set of coordinates. The embedding is non-linear and it preserves the geometry of the original data. For PCA, we tested several values for the dimension of the projected space. Taking the first 100 leading principal coordinates gave the best results. Table 1 summarizes the results from each sensor based on the proposed framework compared to a PCA based dimensionality reduction. Diffusion maps yields very high discrimination results and clearly outperforms PCA. The station’s total score is computed using a majority vote over all sensor based classifiers within the station.

5. Conclusions and Future Work

This paper has introduced a new approach for discrimination between earthquakes and explosions based on non-linear dimensionality reduction methods like diffusion maps. Diffusion maps enabled us to construct a geometric representation of the seismograms that capture the intrinsic structure of the signals. In the pre-processing step, each seismogram was converted into a normalized sonogram. The high dimensional dataset of normalized sonograms was embedded into a low-dimensional space. New seismic events were included in the low-dimensional representation and classified according to the known labels of the training data by applying the k-Nearest Neighbors (k-NN) algorithm. Note that on this post-processing can be applied by support vector machines as well. We evaluated the proposed diffusion-based discrimination algorithm on a data set comprising seismic events in the Dead Sea area with duration magnitudes $\text{Md} \geq 2.5$ that occurred in years 2004-2014. The discrimination results yield a correct discrimination rate that is higher than 90% when applied to test waveforms from two seismic stations on local distances. This demonstrate the feasibility of our method. A next natural step in this framework
is to apply the proposed diffusion-based discrimination algorithms for larger data sets comprising seismic events on regional and teleseismic distances. We also intend to apply kernel based sensor fusion methods (such as [Lindenbaum, 2015], [Lederman, 2015]) for extension of our approach from single stations to seismic networks.

Acknowledgments. This research was supported by the research grant of Pazy Foundation. We are grateful to Dov Zakosky and Batia Reich for providing us with the seismic catalog of Geophysical Institute of Israel.

References


Figure 1. A map of the seismic events and stations.

Figure 2. Histograms of magnitudes for earthquakes (left) and for explosions (right).
Earthquake 200402110815, Md=5.2
Earthquake 200402111936, Md=2.6
Earthquake 201309120120, Md=3.5
Underwater explosion 200410211505, Md=3

Figure 3. Top: Four seismograms: earthquakes with duration magnitudes $Md$ of 5.1, 2.6 and 3.5 and an underwater explosion with $Md = 3$. Bottom: The resulting normalized sonograms.
Figure 4. Diffusion maps embedding of 20 events. Left: the blue points are natural events and the red point is an explosion. Right: the embedded points are colored by their magnitudes. It can be seen that the y-axis captures this intrinsic property.
Figure 5. A 2-dimensional Diffusion representation of the events recorded by three sensors in HRFI. Blue markers- earthquakes. Yellow markers-explosion events.
Figure 6. A 3-dimensional diffusion representation of the events recorded at station HRFI Z sensor. Markers represent the ground truth, color represents the classification results based on 1-NN.
Table 1. Summary of the classification results (in percentages) based on all six sensors, using k-NN (k=1) applied to the 3 or 4 leading diffusion coordinates (# in the brackets) or 100 principal components. The 4-th and 8-th rows present the majority vote score from all sensors.

<table>
<thead>
<tr>
<th>Method</th>
<th>DM (3)</th>
<th>DM (4)</th>
<th>PCA (100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRFI-E</td>
<td>96</td>
<td>96</td>
<td>77</td>
</tr>
<tr>
<td>HRFI-N</td>
<td>96</td>
<td>96</td>
<td>63</td>
</tr>
<tr>
<td>HRFI-Z</td>
<td>98</td>
<td>97</td>
<td>75</td>
</tr>
<tr>
<td>HRFI</td>
<td>98</td>
<td>98</td>
<td>75</td>
</tr>
<tr>
<td>MMLI-E</td>
<td>94</td>
<td>94</td>
<td>70</td>
</tr>
<tr>
<td>MMLI-N</td>
<td>92</td>
<td>92</td>
<td>77</td>
</tr>
<tr>
<td>MMLI-Z</td>
<td>91</td>
<td>93</td>
<td>75</td>
</tr>
<tr>
<td>MMLI</td>
<td>93</td>
<td>93</td>
<td>76</td>
</tr>
</tbody>
</table>