Seismic event identification using diffusion maps

Neta Rabin (1), Yuri Bregman (2), Yochai Ben-Horin (2)

(1) Afeka – Tel Aviv Academic College of Engineering
(2) Soreq Nuclear Research Center, Yavne, Israel

9th Munich Earth Skience School, 17-22 Feb. 2019

Problem formulation

Consider a high-dimensional train dataset $X = \{x_1, x_2, \ldots x_n\}, \quad x_i \in \sim^{m}$

and a function $f : X \to \sim^{D}$ that is defined on the train data.

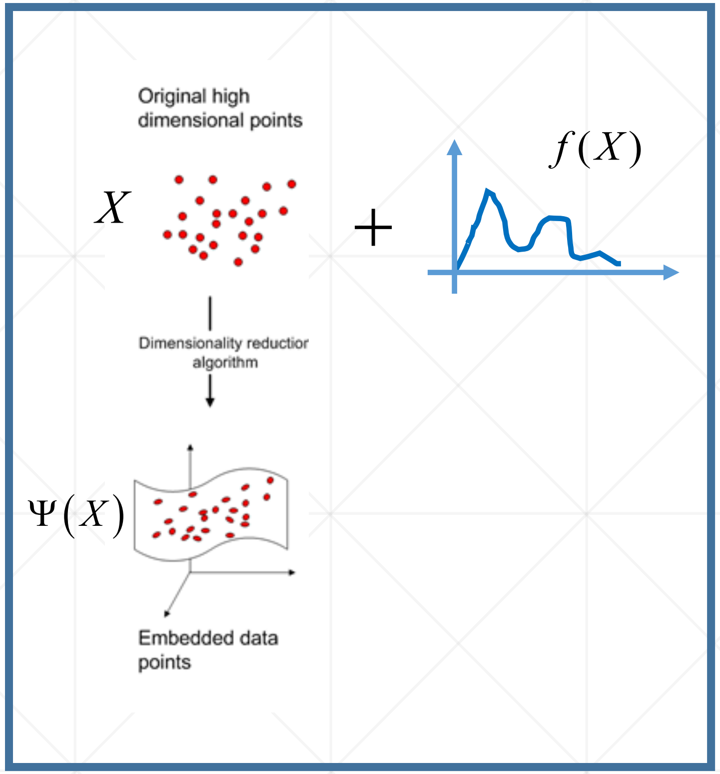
Denote the set of high-dimensional points $\bar{X} \subseteq \sim^{m} \setminus X$ as the test set.

The task is to extend the function f to the points in $\bar{X}$.

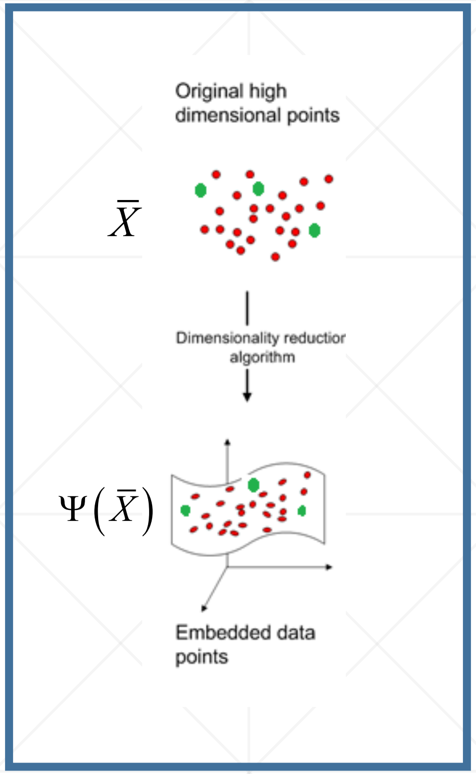
Goal: Find a low-dimensional representation that captures the underlying intrinsic phenomena in X.

Learn and extend the function $f(X)$ from the new representation.

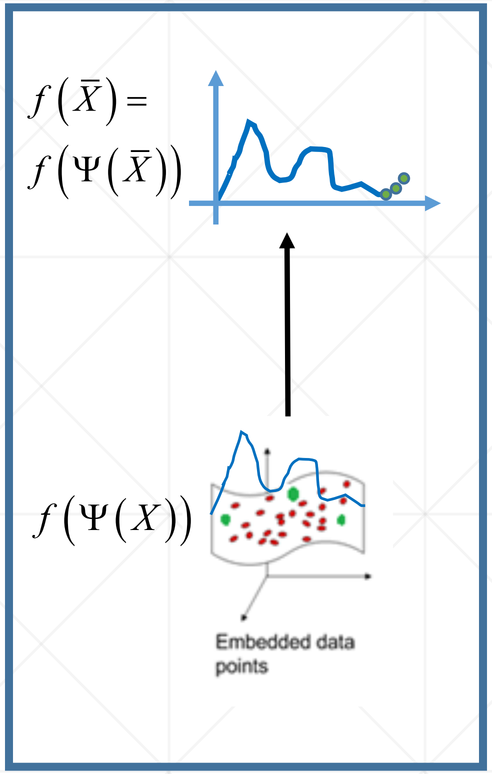
Proposed framework



Embed the data points into a low-dimensional space



Extend the low-dimensional space to include new points



Learn and extend f from the low-dimensional representation

Dimensionality reduction

Motivation

- Data in many real world applications are high dimensional (involve more than two or three dimensions).
- It is hard to interpret high dimensional data.

Goal

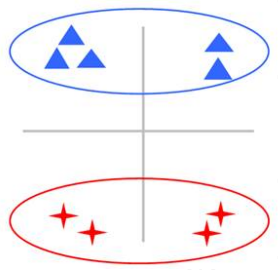
- To simplify the data

Methods

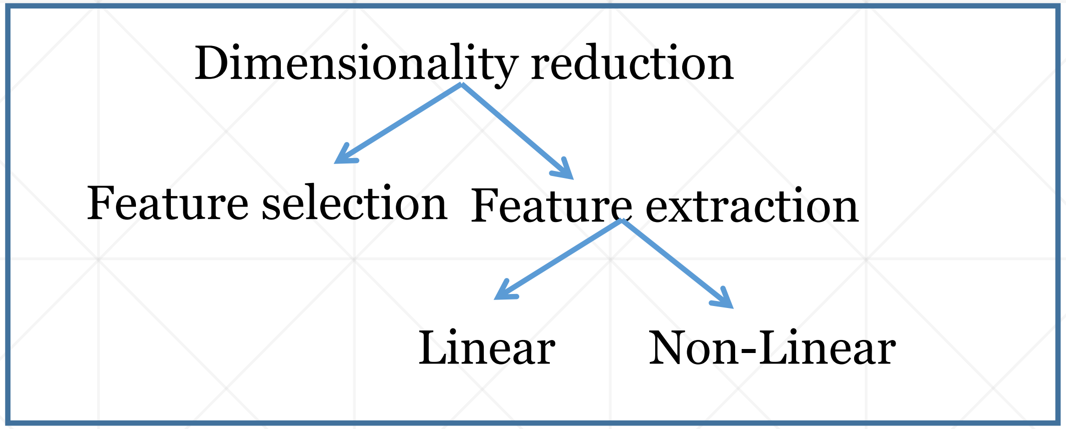
- Embed the data into a low dimensional space, where it is easy to process and analyze.
- The low dimensional space should be faithful to some properties of the original data.

Dimensionality reduction methods

- Feature selection
  - Find a subset of the original variables (features).
  - Algorithms try to identify relevant features

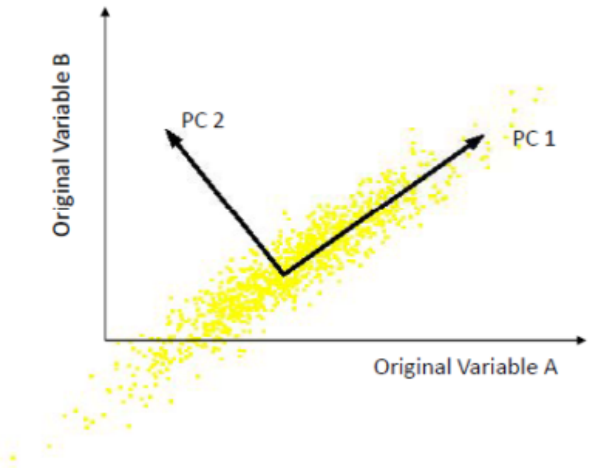


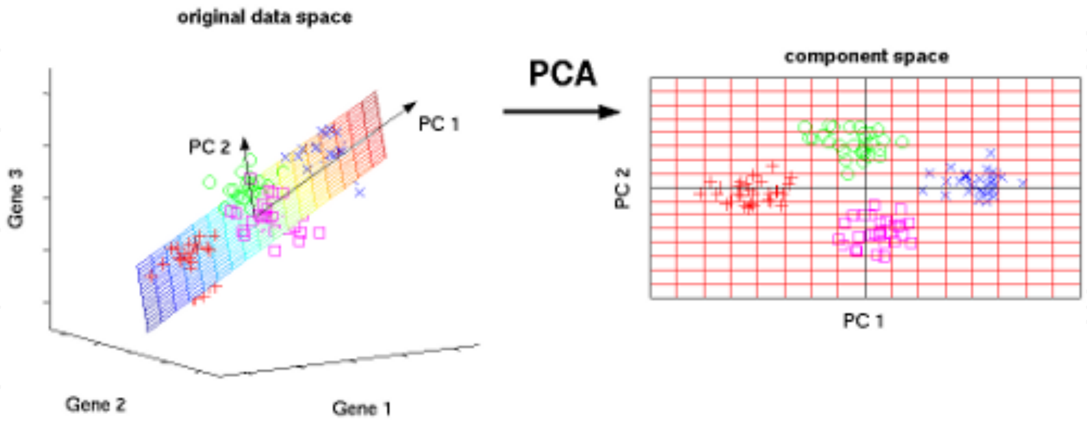
- Feature extraction
  - Transforms data in high-dimensional space to a space of fewer dimensions.



Principal component analysis (PCA)

- A common linear dimensionality reduction method
- Converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called **principal components**.
- The principal components have orthogonal directions of the greatest variance in data.

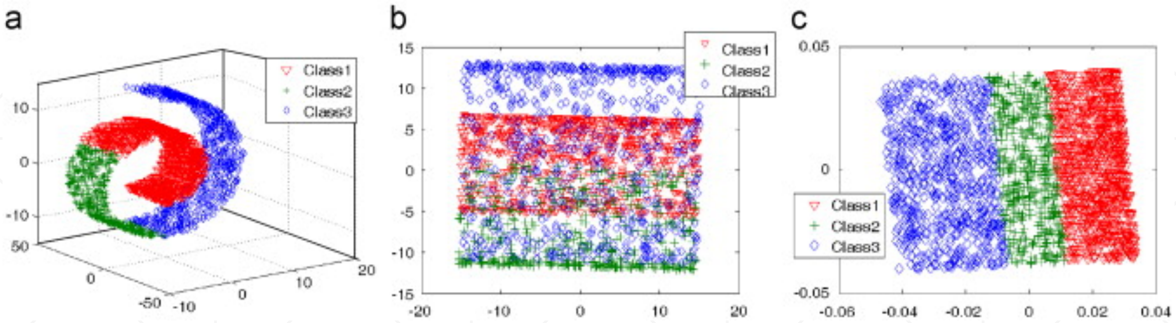




Pictures from: http://www.nlpca.org/pca_principal_component_analysis.html

Nonlinear dimensionality reduction methods

When the data has a nonlinear structure, linear methods like PCA fail



The Swiss roll PCA (linear) Nonlinear

Nonlinear methods, often called manifold learning, usually use graphs, kernel techniques and spectral decomposition to embed the data into a low-dimensional space.

The embedding aim to reflect the intrinsic variables of the data, the values from which the data was produced.

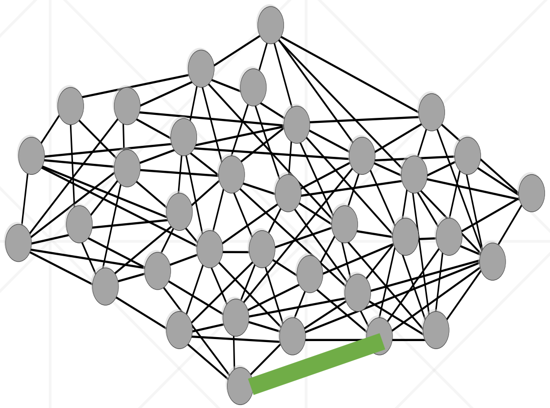
Diffusion maps for nonlinear dimensionality reduction

- A non-linear dimensionality reduction technique
- Construct a global description of the dataset from local similarities
- Uses the eigenfunctions of a Markov matrix
- Preserves the distances of the original data, close points in the original space remain close in the embedded space

R.R. Coifman and S. Lafon Diffusion maps, Applied and Computational Harmonic Analysis, vol. 21, issue 1, pages 5-30, 2006.

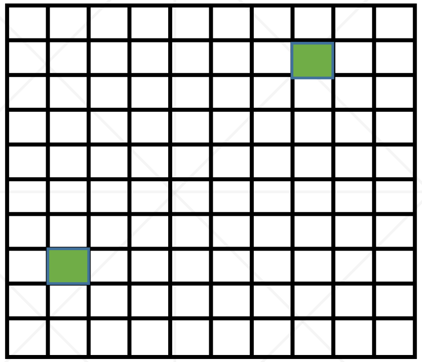
Diffusion maps - schematic description

Graph-structured data



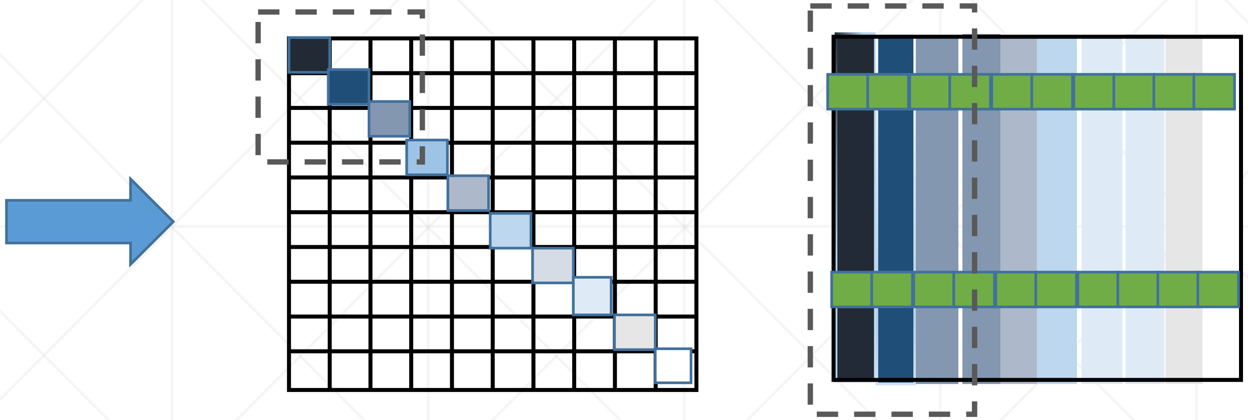
$$G = (X, K)$$

Kernel matrix



$$K_{i,j} = k(x_i, x_j) \propto e^{-\frac{\|x_i - x_j\|^2}{2\varepsilon}}$$

Spectral decomposition and dimensionality reduction



$$\lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots$$

$$\psi_0, \psi_1, \psi_2, \psi_3, \ldots$$

Diffusion maps construction

Input: data points $X = \{x_1, x_2, \ldots x_n\}$ in $\sim^{m}$ **Output:** a low-dimensional embedding $\Psi(X)$ of X

1. Construct a graph $G = \left(X, W\right)$ with a symmetric and positive preserving kernel

   The Gaussian kernel is a common choice $W_{ij} = w\left(x_i, x_j\right) = e^{-\left\|x_i - x_j\right\|^2 / 2\varepsilon}$.

2. Normalize the kernel matrix to be row stochastic: $K_{ij} = k(x_i, x_j) = \dfrac{w(x_i, x_j)}{\sum_j w(x_i, x_j)}$.

3. Compute the spectral decomposition of K:

$\{\lambda_l\}_{l=0}^{n-1}$ are the eigenvalues of K and $\{\phi_l\}_{l=0}^{n-1}$ $\{\psi_l\}_{l=0}^{n-1}$ are the left and right eigenvectors of K.

Diffusion maps construction

K is positive semi-definite matrix, thus, the eigenvalues are real and decay to zero $\lambda_l \to 0$.

Since K is conjugate to a symmetric matrix $A = D^{\frac{1}{2}} K D^{-\frac{1}{2}}$ (with $\{\lambda_l\}_{l=0}^{n-1}$, $\{v_l\}_{l=0}^{n-1}$ orthogonal)

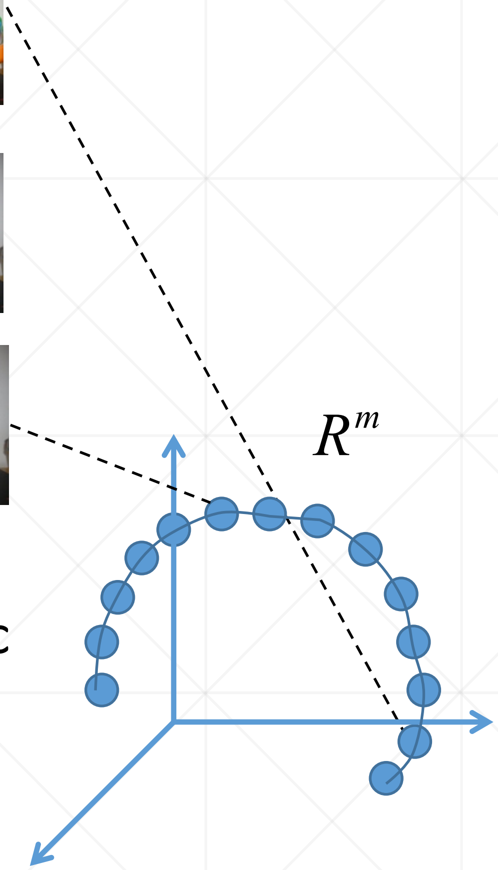
The left and right eigenvectors of K $\phi_l = D^{\frac{1}{2}} v_l$, $\psi_l = D^{-\frac{1}{2}} v_l$ are biorthogonal $\langle \phi_l, \psi_k \rangle = \delta_{lk}$

The spectral decomposition of K is given by $k(x_i, x_j) = \sum_{l \geq 0} \lambda_l \psi_l(x_i) \phi_l(x_j)$.

4. Diffusion map, defined by $\Psi(x_i) = \{\lambda_1 \psi_1(x_i), \lambda_2 \psi_2(x_i), \lambda_3 \psi_3(x_i), \ldots\}$,

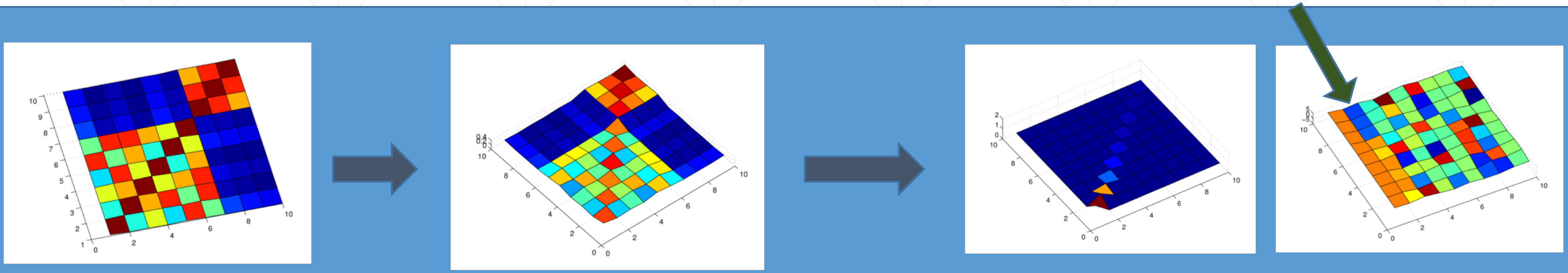
embed the data into Euclidean space.

Toy example



Although each data points is high-dimensional, the intrinsic dimension is 1

1st diffusion maps coordinate



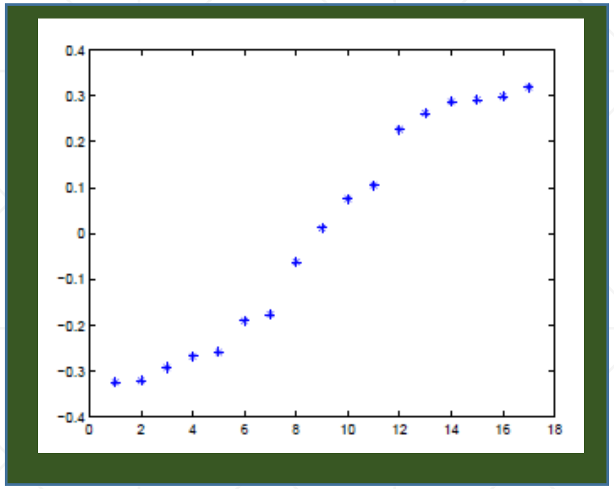
Kernel matrix W

Normalized matrix K

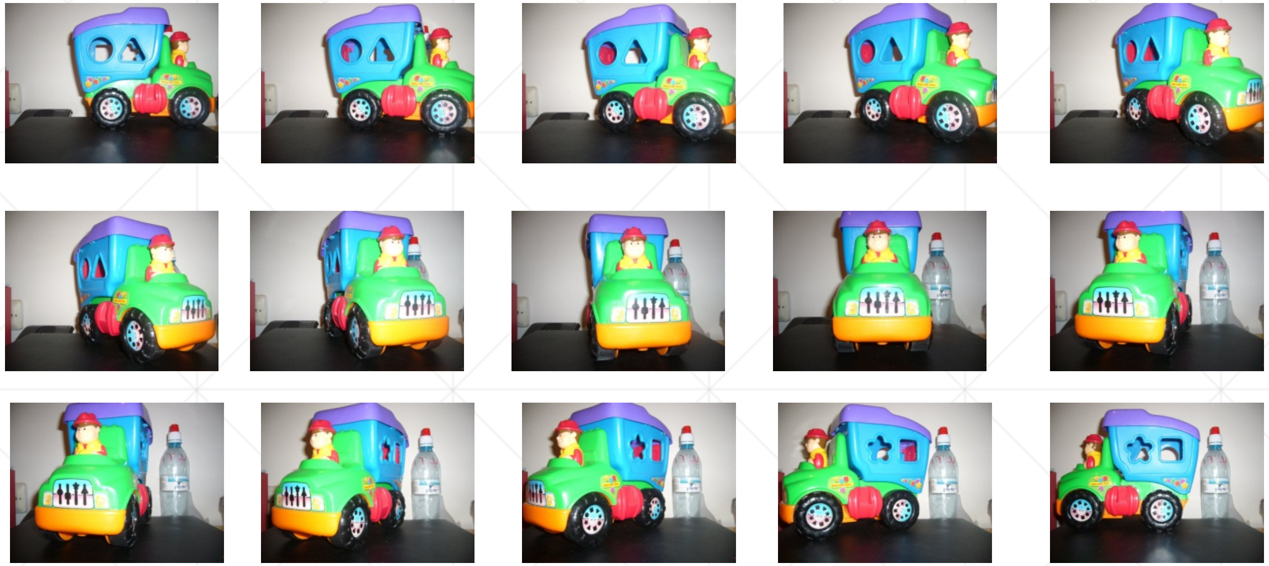
Eigenvalues

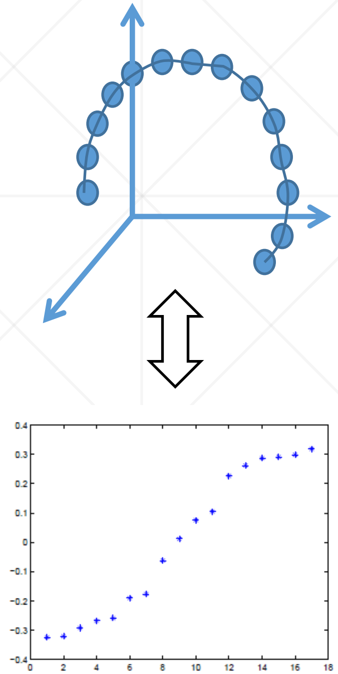
Eigenvectors

The first non-trivial eigenvector, sorted



Toy example organized by DM

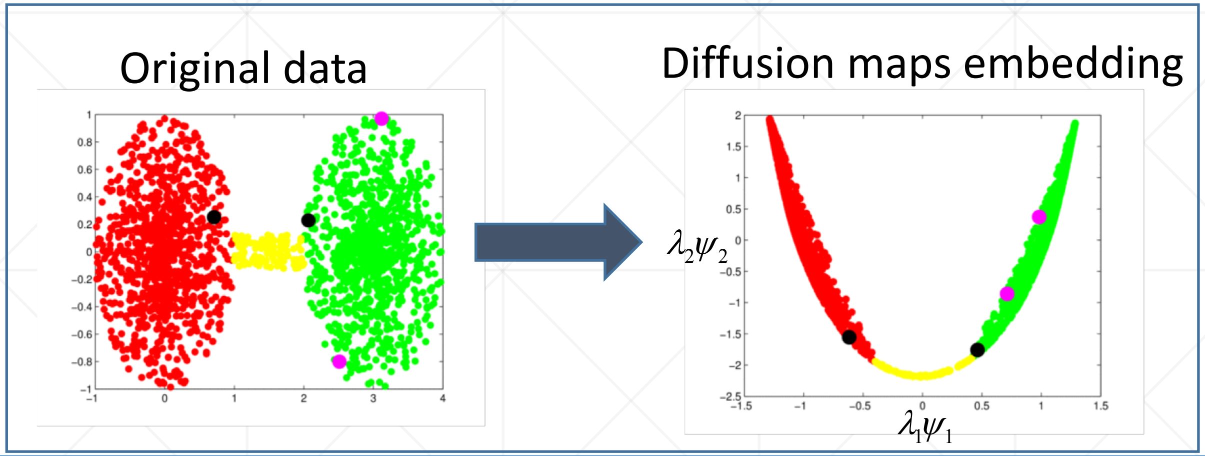




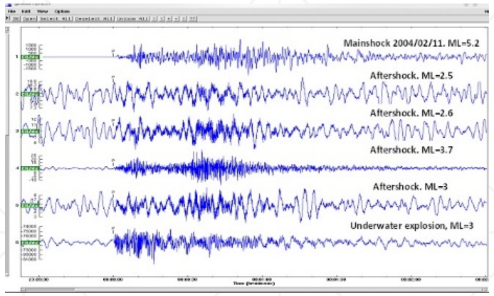
Images sorted according to their value in the first eigenvector

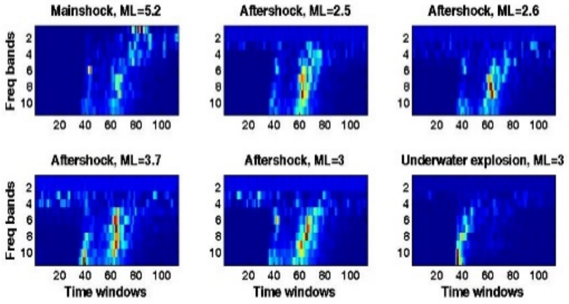
Diffusion distances

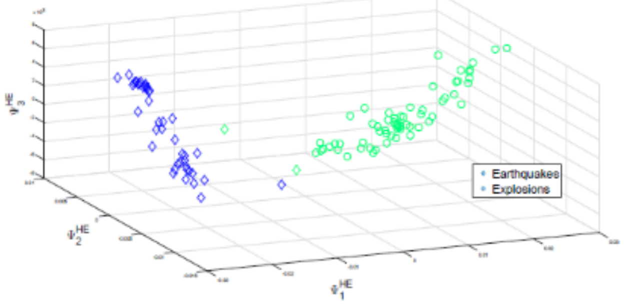
- The diffusion distance between two data points is $D(x_i, x_j)^2 = \sum_{x_l \in \Gamma} \frac{(k(x_i, x_l) - k(x_l, x_j))^2}{\varphi_0(x_l)}$
- Substituting $k(x_i, x_l) = \sum_{l \geq 0} \lambda_l \psi_l(x_i) \phi_l(x_l)$ yields $D(x_i, x_j)^2 = \sum_{l \geq 1} \lambda_l (\psi_l(x_i) - \psi_l(x_j))^2$
- This metric preserves pairwise distances and embeds the data into Euclidean space.



Earthquake-explosion discrimination







Method	DM (3)	DM (4)	PCA (100)
HRFI-E	96	96	77
HRFI-N	96	96	63
HRFI-Z	98	97	75
HRFI	98	98	75
MMLI-E	94	94	70
MMLI-N	92	92	77
MMLI-Z	91	93	75
MMLI	93	93	76

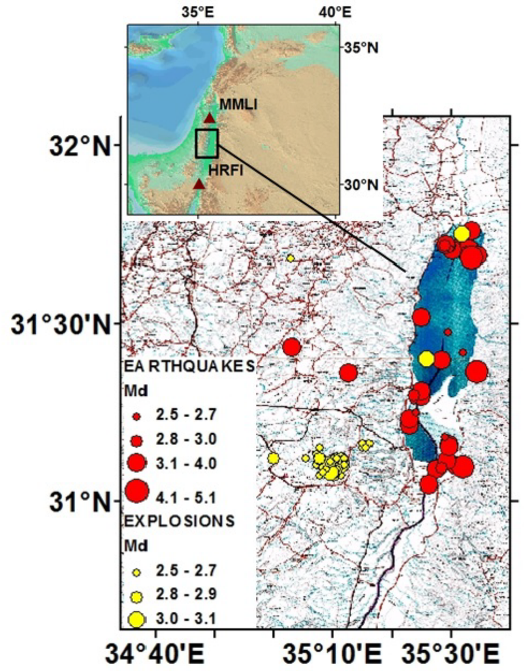
Outperforms PCA

Input: seismic waveforms
Preprocessing: STFT
Task: Earthquake-explosion discrimination

Earthquake-explosion discrimination using diffusion maps,
N. Rabin, Y. Bregman, O. Lindenbaum, Y. Ben-Horin and A Averbuch,Geophysical Journal International 207 (3), 1484-1492, 2016.

Dataset

- The dataset includes 44 earthquakes and 62 explosions.
- The dataset was recorded at two different stations.
- The collected waveforms occurred in the Dead Sea area between 2004 and 2014.
- The duration magnitudes are Md>2.5.
- The waveforms were independently analyzed by the Geophysical Institute of Israel (GII), thus reliably labeled (earthquake or explosion).



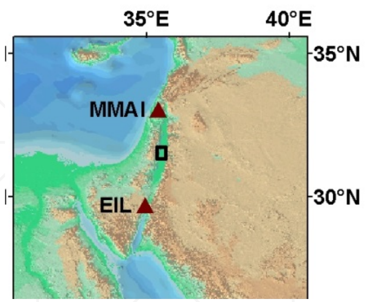
Preprocessing – sonogram computation

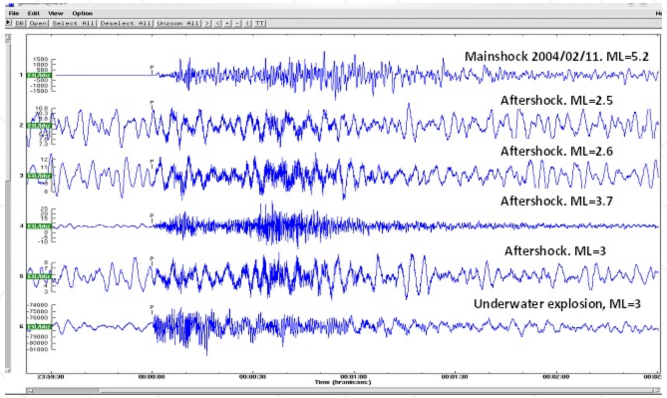
- The waveform is aligned by the estimated P phase onsets.
- Overlapping sliding windows of length 2^L samples are created.
- Sums of power spectral densities for logarithmically scaled frequency bands.
- The bands are slightly overlapping (by one bin).
- Tapering Hann windows before FFT are applied.
- The values are normalized to overcome magnitude differences.

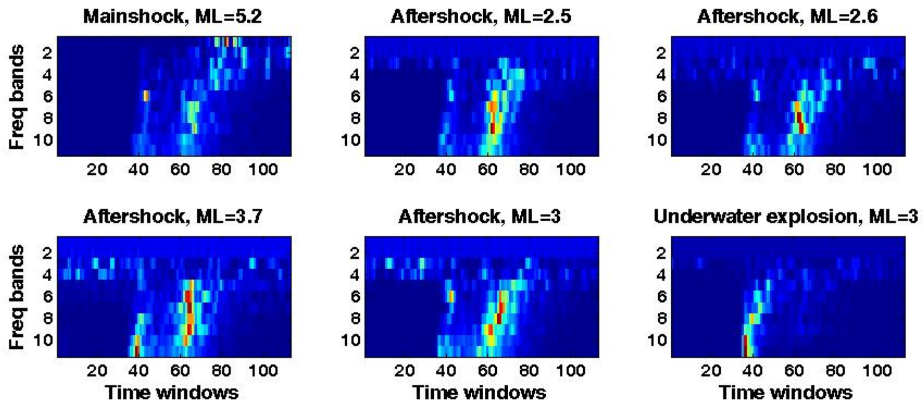
Preprocessing example

Six events that occurred in the Dead Sea recorded by a station in Eilat, EIL, ~250 km from the dead sea

- One main earthquake shock (Magnitude 5.2)
- Four aftershocks that are occurred after the main shock
- One underwater explosion

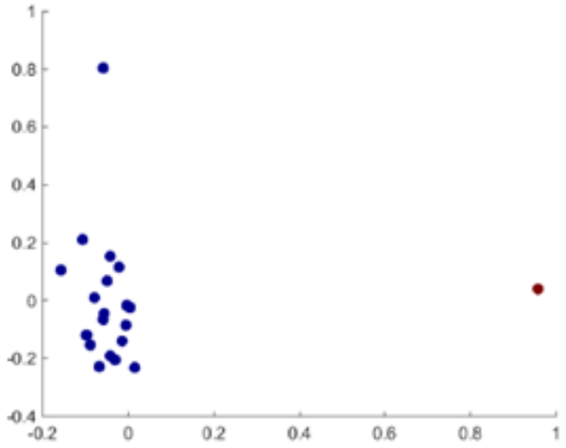




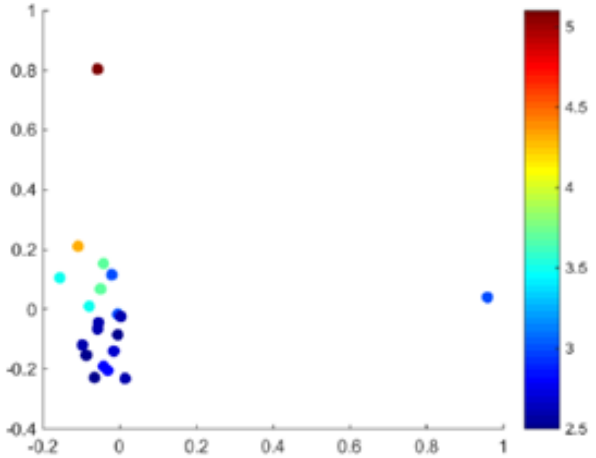


Diffusion maps embedding of 20 events.

- A subset of 20 events from the northern Dead Sea area.
- 19 events are earthquakes (duration magnitudes: 2.5 - 5.1).
- One event is an explosion of magnitude 3.

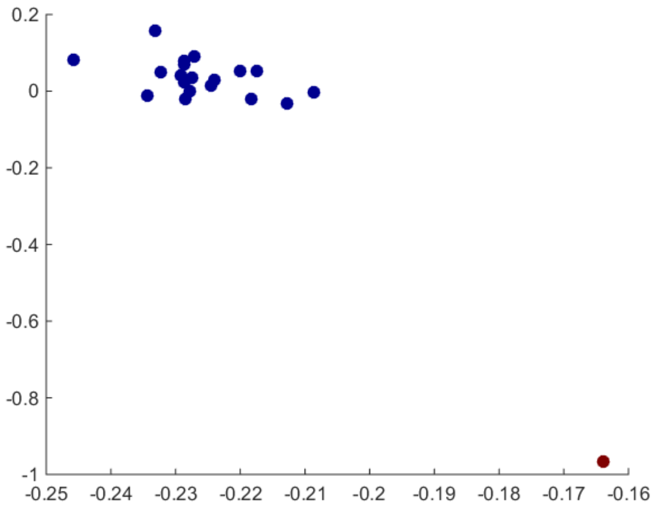


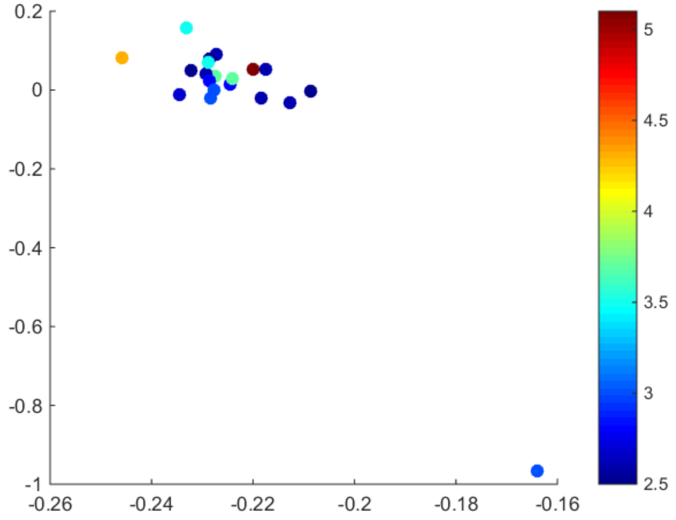
Left: earthquakes (blue) and an explosion (red)



Right: embedded points colored by their magnitudes. The y-axis captures this intrinsic property.

Comparison to linear (PCA) projection

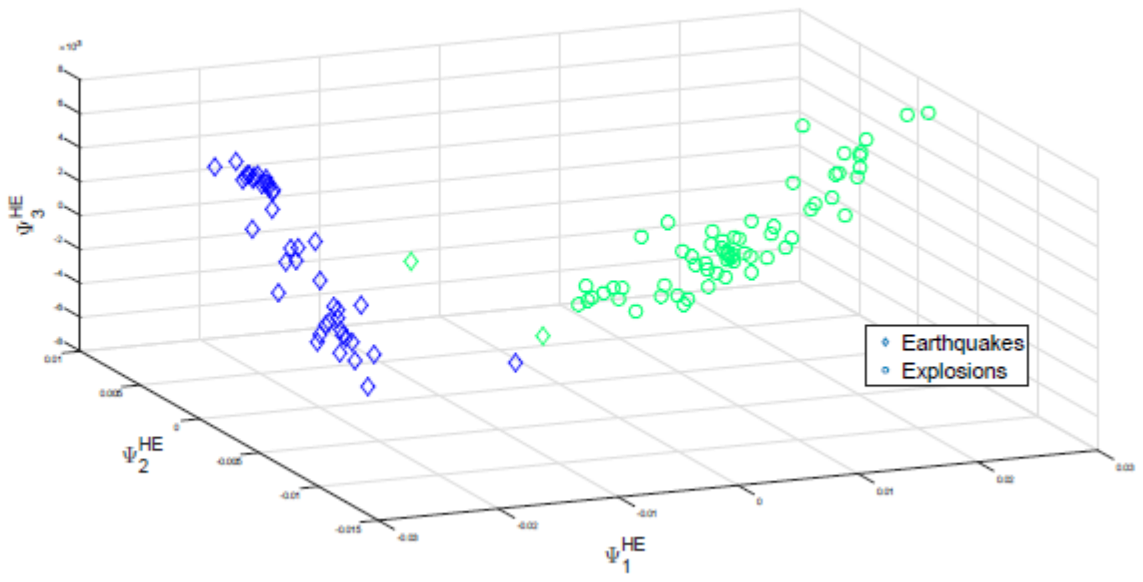




Left: earthquakes (blue) and an explosion (red)

Right: projected points colored by their magnitudes.

Diffusion maps for event discrimination



A 3-dimensional diffusion maps representation of the events recorded by the vertical channel of HRFI.

Markers represent true label, color represents the classification results based on 1-NN.

Leave one out Cross Validation

Method	DM (3)	DM (4)	PCA (100)
HRFI-E	96	96	77
HRFI-N	96	96	63
HRFI-Z	98	97	75
HRFI	98	98	75
MMLI-E	94	94	70
MMLI-N	92	92	77
MMLI-Z	91	93	75
MMLI	93	93	76

Summary of the classification results (in percentages) based on all six channels, using 1-NN applied to the 3 or 4 leading diffusion coordinates (# in the brackets) or 100 principal components.

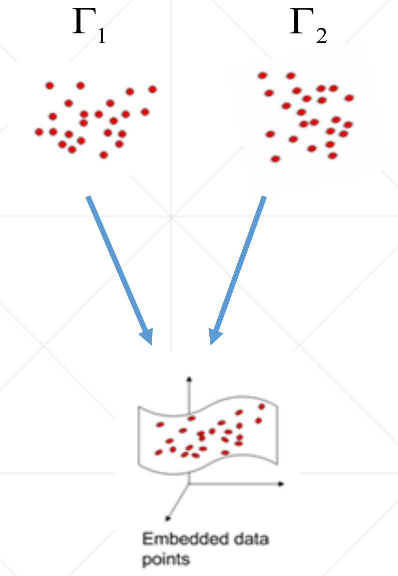
The 4-th and 8-th rows present the majority vote score from all sensors.

Methods: multi-view diffusion maps

Problem formulation:

Consider two aligned datasets $\Gamma_1 = \{x_1, x_2, \dots x_N\}, \quad x_i \in R^{n_1}$ and $\Gamma_2 = \{y_1, y_2, \dots y_N\}, \quad y_i \in R^{n_2}$ that hold the same number of observations.

The data points x_i and y_i describe the same entity from different views (sensors)



Goal: construct a low-dimensional representation that captures the mutual information from the two views.

* O. Lindenbaum, A. Yeredor, M. Salhov, A. Averbuch, Multiview diffusion maps, arXiv preprint, 2015
* R. R. Lederman, R. Talmon, Learning the geometry of common latent variables using alternating-diffusion, Applied and Computational Harmonic Analysis, vol. 44, issue 3, pp. 509-536, 2018.

Construction

- The points in Γ_1, Γ_2 are organized as a graph with $K_1,\ K_2$ as the edge matrices

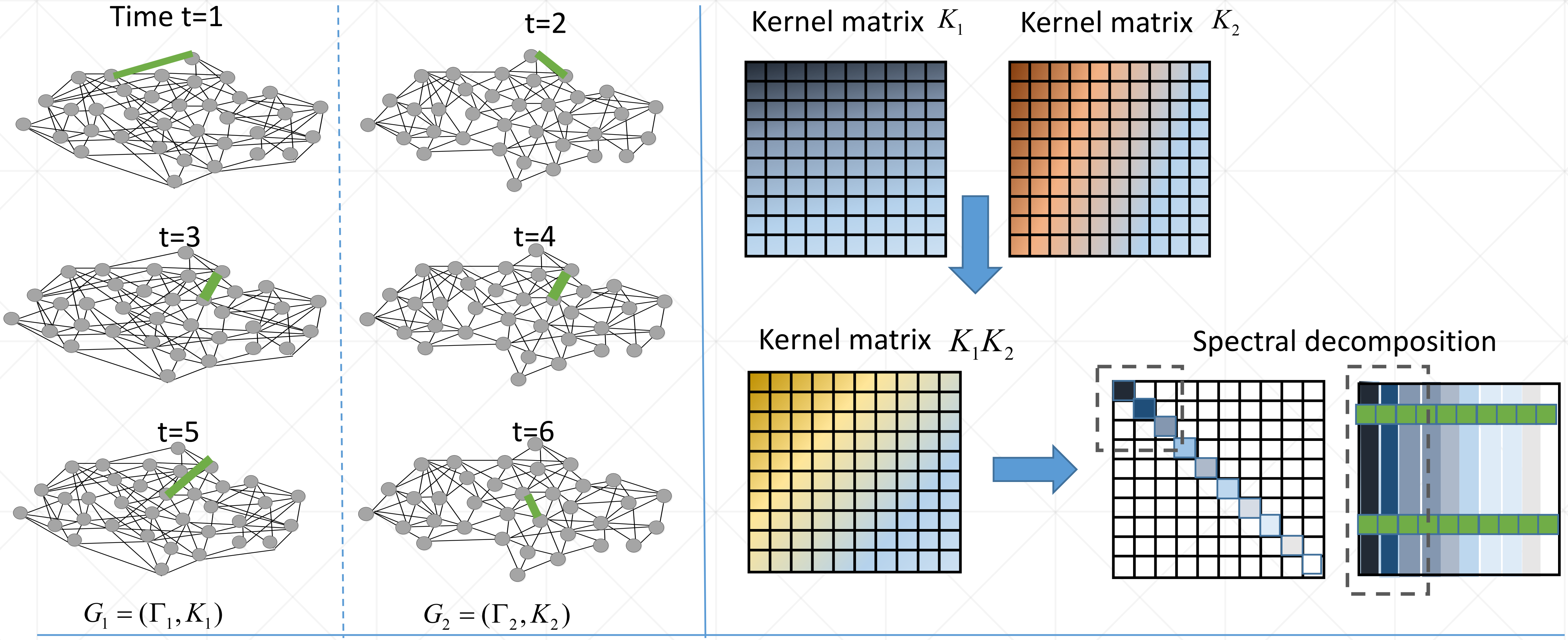
- The multi-view kernel $K = K_1 K_2$ is constructed

- The probabilities for the odd jumping steps are given by K_1
- The probabilities for the even jumping steps are given by K_2

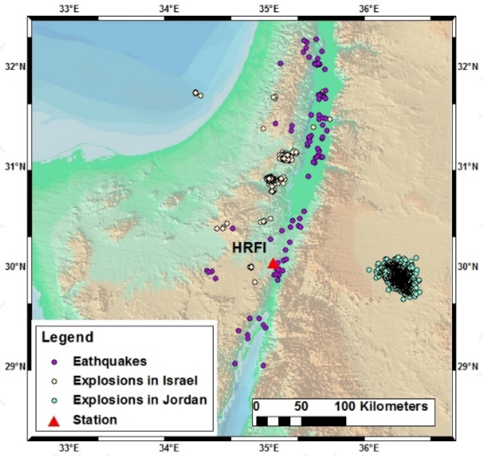
- Only edges that contain mutual information of both diffusion processes have a meaningful value in K

- The spectral decomposition may be computed to form a common embedding

Schematic description



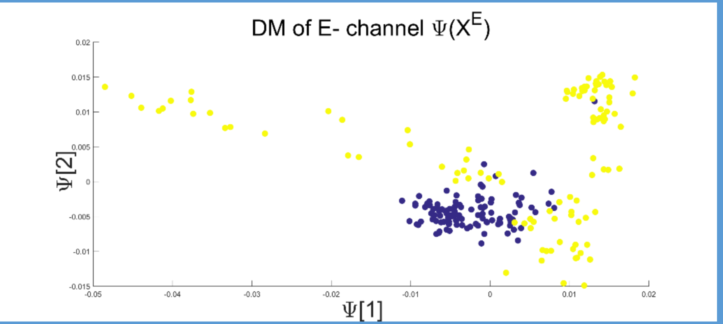
Application: multi-view kernels for modeling seismic events (earthquake explosion discrimination)



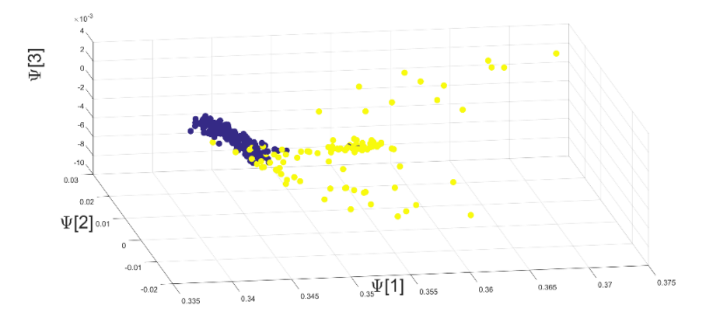
Input: Seismic waveforms of Earthquakes (~100)
and explosions (~2000)

Data from 3 channels (Z N E)

Diffusion maps (single channel)

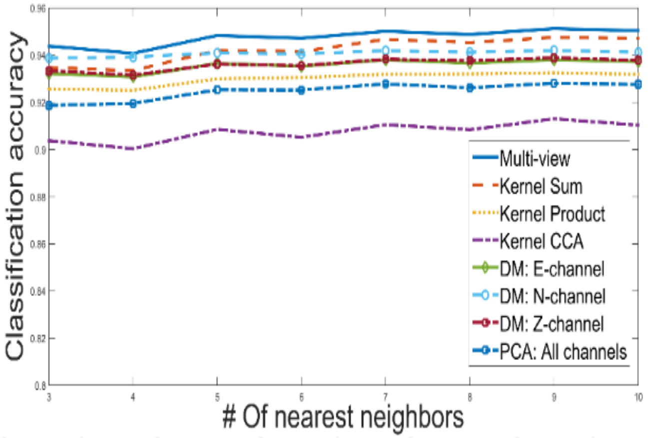


Multi-view diffusion maps (3 chnls.)



EXP – in blue, EQ – in yellow

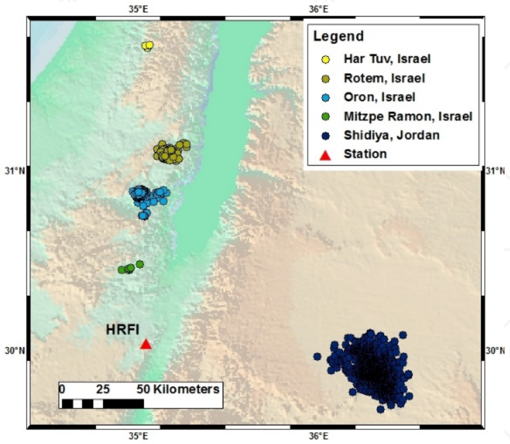
Classification results



Multi-view yields the best results

O. Lindenbaum, Y. Bregman, **N. Rabin** and A. Averbuch, Multiview Kernels for Low-Dimensional Modeling of Seismic Events, IEEE Transactions on Geoscience and Remote Sensing, vol. 56, issue 6, pp. 3300-3310, 2018.

Multi-view kernels for modeling seismic events (Quarry classification)

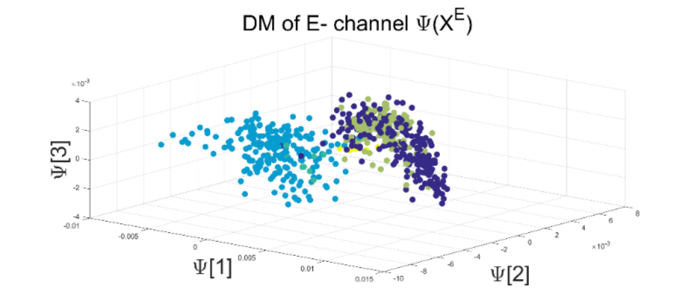


Input: Seismic waveforms of 602 explosions from 5 quarries.

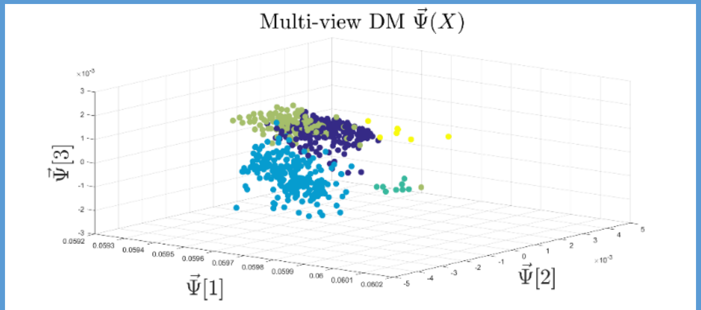
Data from 3 channels (Z N E)

Task: Classify quarry by source.

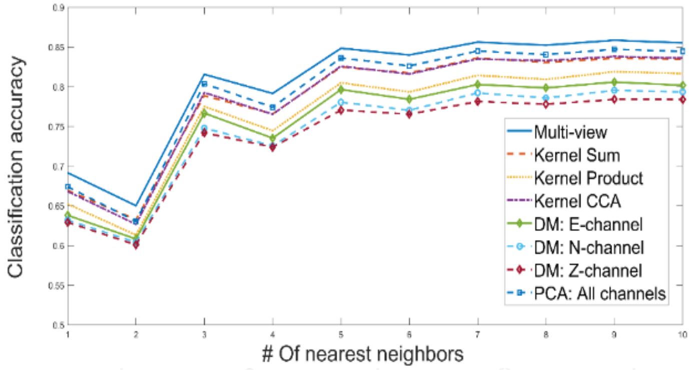
Diffusion maps (single channel)



Multi-view diffusion maps (3 chnls.)



Classification results



Multi-view yields the best results

O. Lindenbaum, Y. Bregman, **N. Rabin** and A. Averbuch, Multiview Kernels for Low-Dimensional Modeling of Seismic Events, IEEE Transactions on Geoscience and Remote Sensing, vol. 56, issue 6, pp. 3300-3310, 2018.

Summary

- Data driven kernel methods enhance capabilities in seismic signal processing.
- They overcome the limitations of traditional parameter-based discrimination techniques.
- Kernel based data fusion is a natural extension for multi-channel and multi-station processing.

Neta.rabin@gmail.com