Diffusion Maps for Changing Data

Matthew J. Hirn

September 3, 2013

Collaborators

- Simon Adar, Tel Aviv University
- Eyal Ben Dor, Tel Aviv University
- Ronald Coifman, Yale University
- Roy Lederman, Yale University
- Nicholas Marshall, Clarkson University
- Yoel Shkolnisky, Tel Aviv University

Overview

Embedding of closed curve

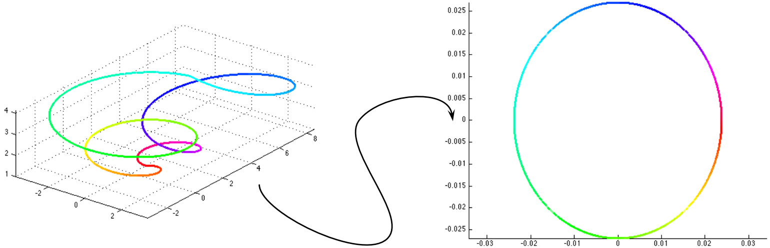
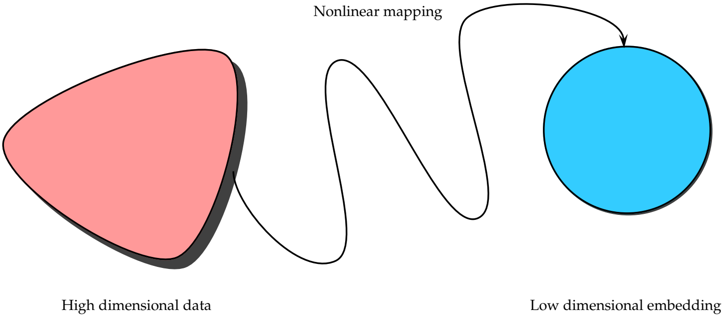


Figure: Left: A closed, non-self-intersecting curve in 3 dimensions. Right: Its embedding as a circle.

Cartoon


Nonlinear mapping
High dimensional data
Low dimensional embedding

Heat equation

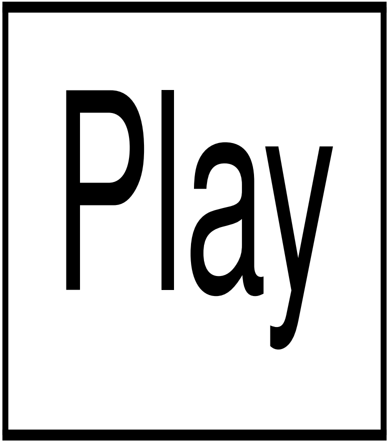
- Let $(\mathcal{M}, g)$ be a compact Riemannian manifold.
- Solve heat equation for $u(x,t)$, $x, y \in \mathcal{M}$, $t \in [0,T]$, with initial condition $\delta_y(x)$,

$$\begin{aligned} \frac{\partial u}{\partial t} &= \Delta_x u \\ u(x,0) &= \delta_y(x) \end{aligned} \tag{1}$$

- The solution to (1) is given by the heat kernel $K(t,x,y)$.
- To compare two points $x, y \in \mathcal{M}$, compute their diffusion distance:

$$D^{(t)}(x,y)^2 = \int_{\mathcal{M}} (K(t,x,z) - K(t,y,z))^2 \, dV(z).$$

Heat flow on sphere



Data driven approximation

- Suppose our data lies (approximately) on a compact Riemannian manifold $\mathcal{M}$.
- The data consists of samples $X = \{x_1, \ldots, x_n\} \subset \mathcal{M} \subset \mathbb{R}^D$.
- To approximate heat flow use a random walk $P_{i,j} = p(x_i, x_j)$ on X, where $p(x_i, x_j)$ is the probability of walking from x_i to x_j in one step.
- P must be local. That is, only connect points that are very close on $\mathcal{M}$, i.e., those that we can empirically have very high confidence that they are similar data points. Additionally, heat diffuses smoothly, so our random walk should walk "approximately smoothly" (i.e., not jump from one area of the manifold to another).
- Powers P^t run the random walk forward, which in turn approximates heat flow up to time t.

Diffusion maps [Coifman and Lafon, 2006]

- The diffusion map is derived from the eigenvectors and eigenvalues of the random walk P.
- Let $\{\psi_1, \ldots, \psi_n\}$ be the eigenvectors of P with corresponding eigenvalues $1 = \lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n| \geq 0$.
- The diffusion map at time t (i.e. after running the random walk forward t steps) is $\Psi^{(t)} : X \to \mathbb{R}^d$, where $d \leq n$ and

$$\Psi^{(t)}(x_i) = (\lambda_1^t \psi_1(x_i), \ldots, \lambda_d^t \psi_d(x_i)).$$

- The diffusion map gives a (low dimensional) embedding of the data that characterizes its geometry according to the diffusion distance:

$$\|\Psi^{(t)}(x_i) - \Psi^{(t)}(x_j)\|_{\mathbb{R}^d} \approx D^{(t)}(x_i, x_j).$$

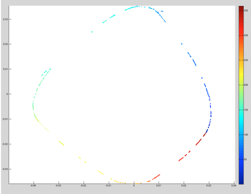
Vehicle trajectories

Play

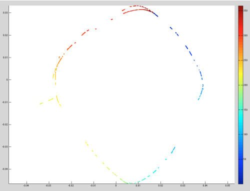
(a) BRM

Play

(b) BTR



(c) BRM



(d) BTR

Figure: Embedding of the videos. The color corresponds to the angle on the circle at which the vehicle is located.

Separating the trajectories

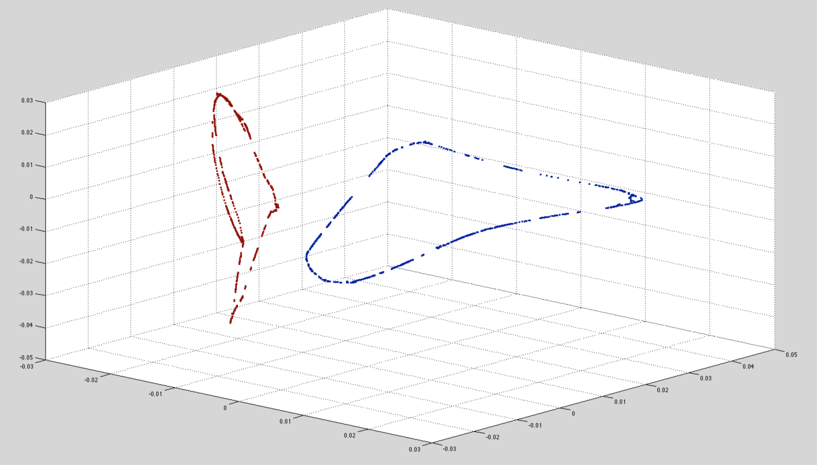


Figure: Embedding of BRM and BTR.

Overview

Time decoupled heat equation

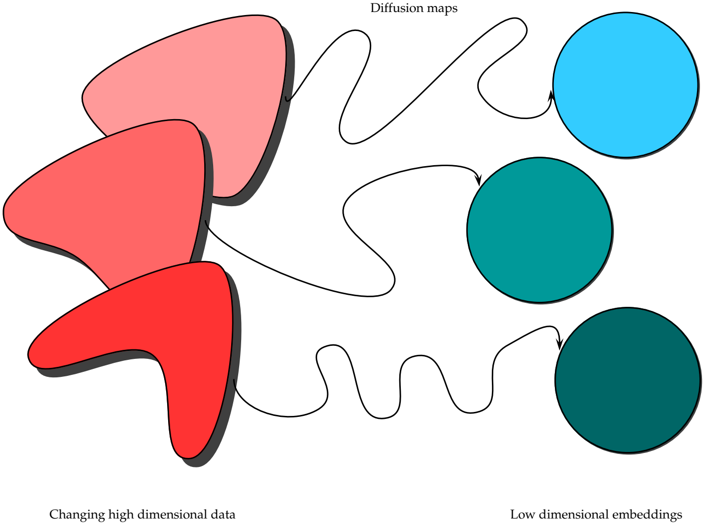
- Let $(\mathcal{M}, g(t))$ be a compact Riemannian manifold with a family of Riemannian metrics $g(t)$, $t \in [0, T]$.
- For a given time $s \in [0, T]$, solve the heat equation for $u(x, t)$, $x, y \in \mathcal{M}$, $t \in [0, T]$, with initial condition $\delta_y(x)$,

$$\begin{aligned} \frac{\partial u}{\partial t} &= \Delta_{g(s)} u \\ u(x, 0) &= \delta_y(x) \end{aligned} \tag{2}$$

- The solution to (2) is given by the heat kernel $K_s(t, x, y)$.
- To compare two points $x \in (\mathcal{M}, g(r))$ and $y \in (\mathcal{M}, g(s))$, we compute a generalized diffusion distance:

$$D^{(t)}(x, y)^2 = \int_{\mathcal{M}} (K_r(t, x, z) - K_s(t, y, z))^2 \, dV(z, 0).$$

Changing data



Diffusion maps revisited

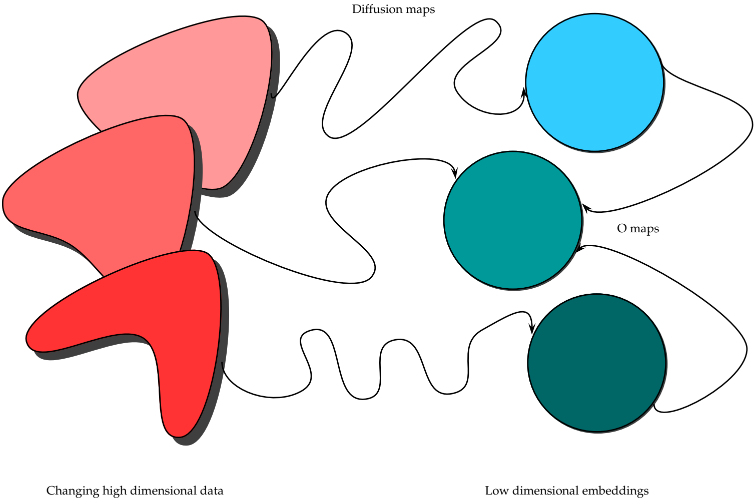
- Data consists of samples

$$X_s = \{x_{s,1}, \ldots, x_{s,n}\} \subset \varphi_s(\mathcal{M}, g(s)) \subset \mathbb{R}^{D_s}.$$

- For each time s, obtain a diffusion map $\Psi_s^{(t)} : X_s \to \mathbb{R}^{d_s}$.
- For each time s one can explicitly define a map $O_s : \mathbb{R}^{d_s} \to \mathbb{R}^d$ such that

$$\|O_r \Psi_r^{(t)}(x_{r,i}) - O_s \Psi_s^{(t)}(x_{s,j})\|_{\mathbb{R}^d} \approx D^{(t)}(x_{r,i}, x_{s,j}).$$

Changing data



Hyperspectral Imagery

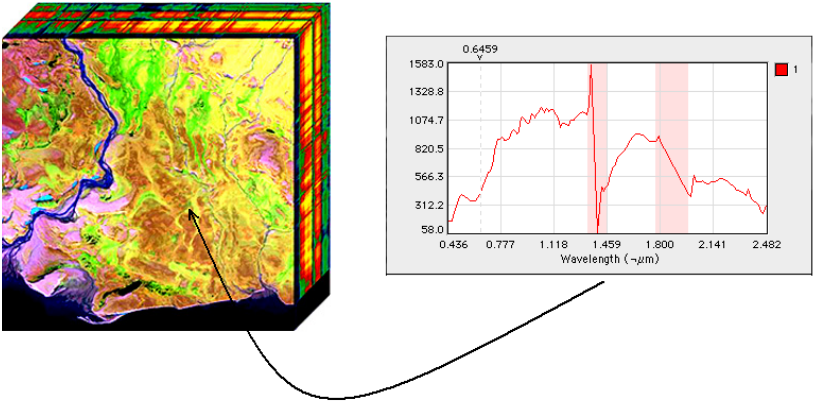


Figure: Hyperspectral image with a pixel signature.

Spectral signatures

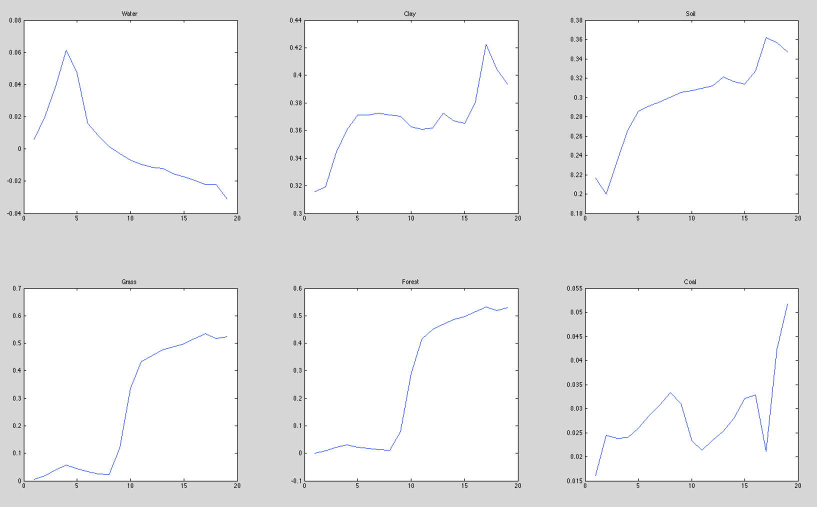


Figure: Spectral signatures of various materials.

Sokolov Mine

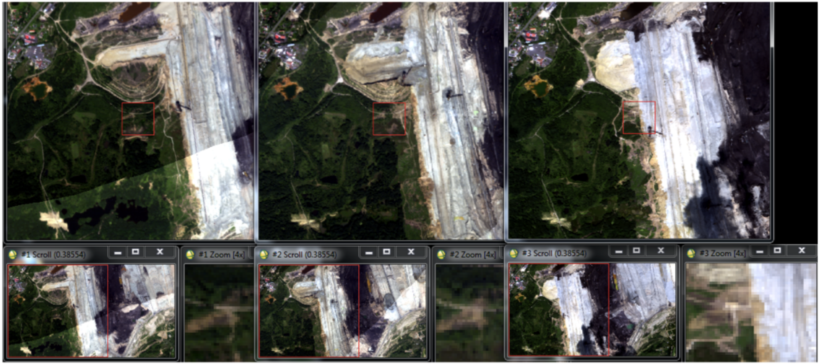
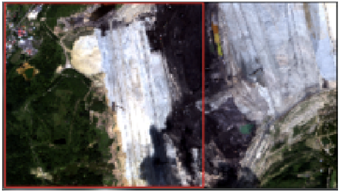


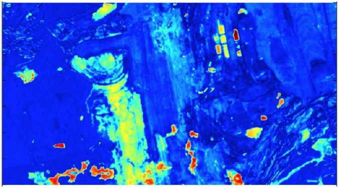
Figure: Sokolov Mine in 2009 and 2010 (during the day, visible range, HyMap sensor), as well as 2011 (during the day, visible range, AHS sensor)

Change Detection using Diffusion Maps

(a) Sokolov Mine 2009 / Day / Visible / HyMap sensor



(b) Sokolov Mine 2011 / Day / Visible / AHS sensor



(c) Change Map using Diffusion Maps

Sokolov Mine

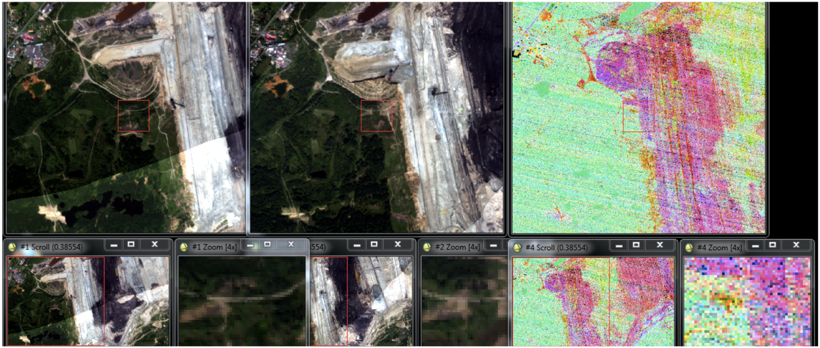
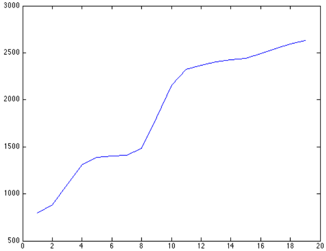
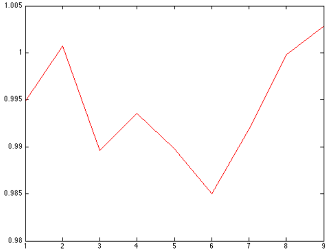


Figure: Sokolov Mine in 2009 and 2010 (during the day, visible range, HyMap sensor), as well as 2011 (at night, thermal range, AHS sensor)

Sample Spectra



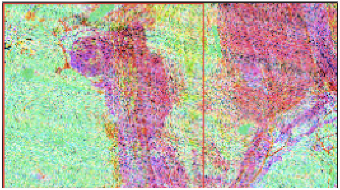
(a) Spectral signature of a pixel, 2009, during the day, visible range



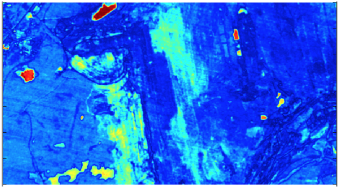
(b) Spectral signature of the same pixel, 2011, at night, thermal range

Change Detection using Diffusion Maps

(c) Sokolov Mine 2009 / Day / Visible / HyMap sensor

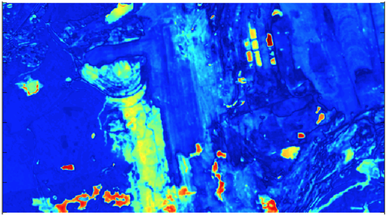


(d) Sokolov Mine 2011 / Night / Thermal / AHS sensor

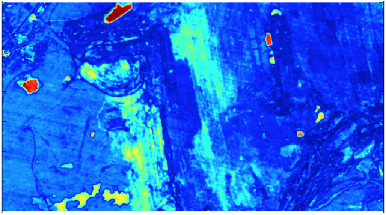


(e) Change Map using Diffusion Maps

Comparison of Change Maps



(a) 2009 / Day / Visible / HyMap sensor vs. 2011 / Day / Visible / AHS sensor



(b) 2009 / Day / Visible / HyMap sensor vs. 2011 / Night / Thermal / AHS sensor

Diffusion Embedding of the Standard Map

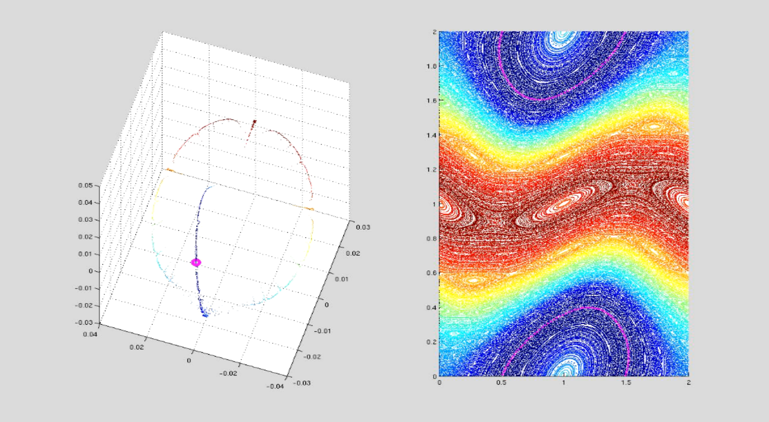


Figure: Standard map on right; diffusion embedding on left. The orbit of a particular color on the right corresponds to the embedded point of the same color on the left.

Diffusion Embedding as K Changes

Figure: Standard map on right; diffusion embedding on left.

Common Embedding

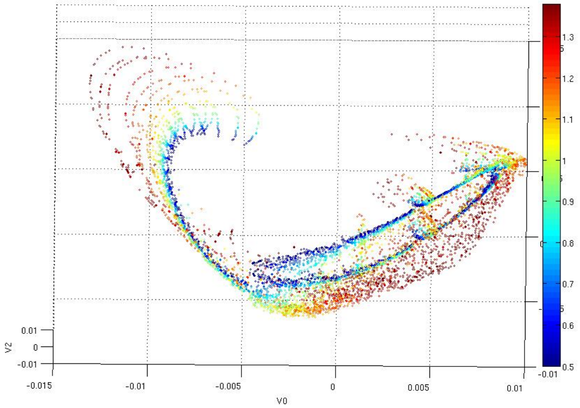


Figure: Diffusion embeddings of the standard map for various K all mapped into the same Euclidean space. The color of the points corresponds to the value of K (see color bar on right).

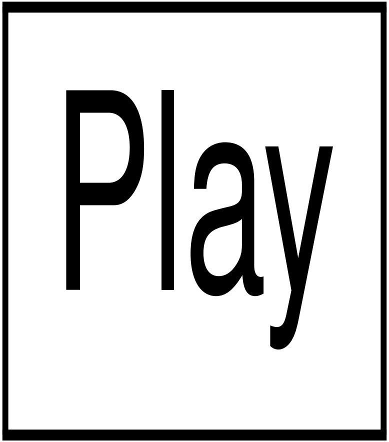
Time coupled heat equation

- Let $(\mathcal{M}, g(t))$ be a compact Riemannian manifold with a family of Riemannian metrics $g(t)$, $t \in [0, T]$.
- For a given time $s \in [0, T]$, solve the heat equation (with time dependent Laplacian) for $u(x, t)$, $x, y \in \mathcal{M}$, $t \in [s, T]$, with initial condition $\delta_y(x)$,

$$\begin{aligned} \frac{\partial u}{\partial t} &= \Delta_{g(t)} u \\ u(x, s) &= \delta_y(x) \end{aligned} \tag{3}$$

- The solution to (3) is given by the heat kernel $Z(s, t, x, y)$.

Diffusion on evolving sphere



Data driven approximation and diffusion map

- Data consists of:
  1. Manifold samples $X = \{x_1, \ldots, x_n\} \subset \mathcal{M} \subset \mathbb{R}^D$.
  2. Time samples $0 = t_0 < t_1 < \cdots < t_m = t \leq T$.
- For each time t_j, approximate the heat flow over X from time t_j to t_{j+1} with one step of a random walk P_{t_j}.
- The diffusion over the time evolving manifold up to time t is then approximated by:

$$P^{(t)} = P_{t_m} P_{t_{m-1}} \cdots P_{t_0}$$

- Compute a diffusion map from the eigenvectors and eigenvalues of $P^{(t)}$.

Stability of embedding

(a) Evolving manifold

(b) Top: Time coupled; Bottom: Time-decoupled

Thank you