Diffusion Based Manifold Learning
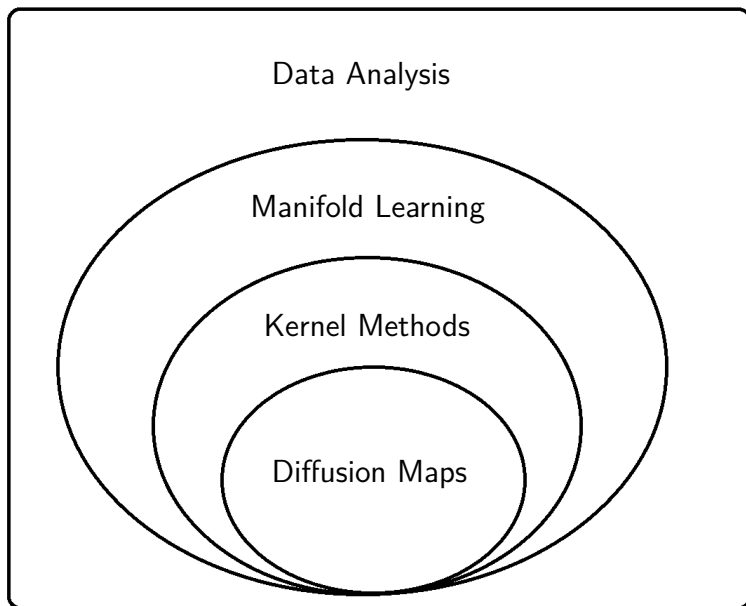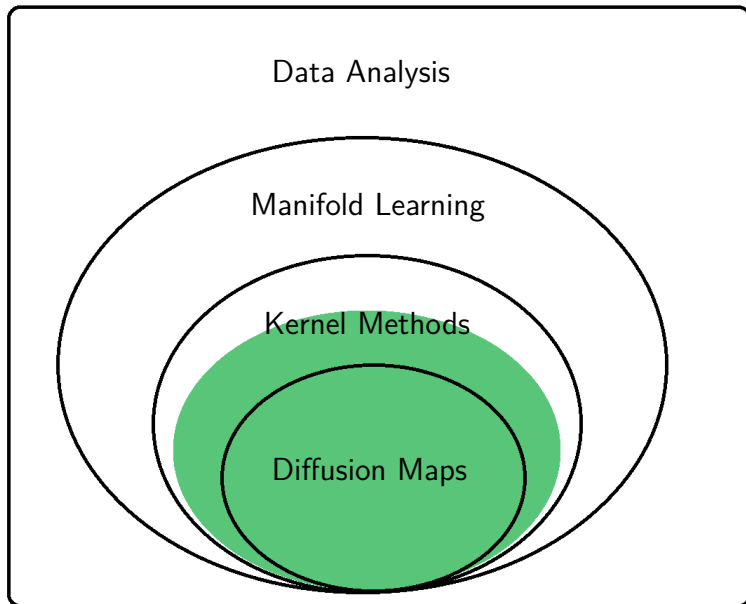
Matthew Hirn and Guy Wolf

October 23, 2013

Data Analysis Scope

Data Analysis

Manifold Learning

Kernel Methods

Diffusion Maps

Data Analysis Scope

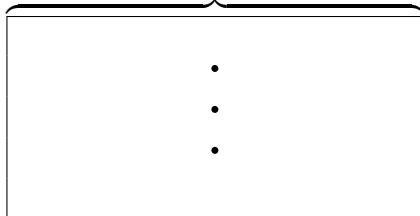# Overview

Overview

Manifold learning
Why use manifolds?

availability of data
⇓
many observable parameters
⇓
high-dimensional recorded data
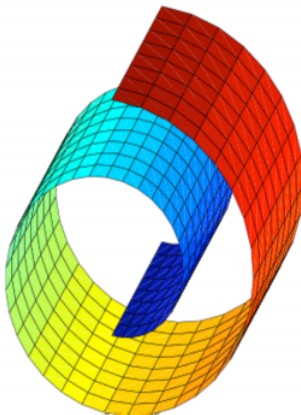
· · ·

- Data contain dependencies & redundancies
- Observable space = non-linear mapping of few underlying factors
- Underlying locally low-dimensional geometry in high-dimensional ambient space

Manifold learning

The goal

Manifold learning
Kernel methods

Lips

(a) One (b) Five
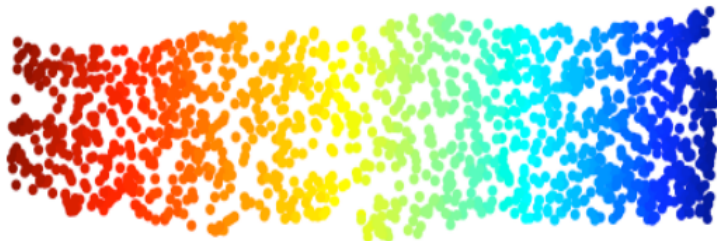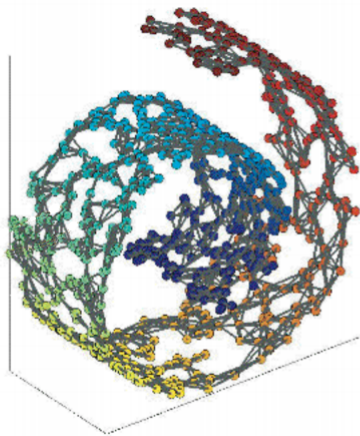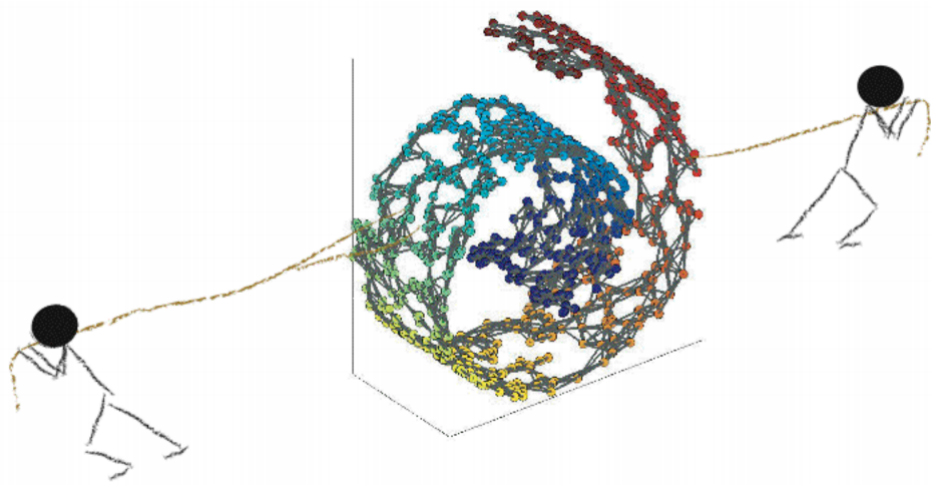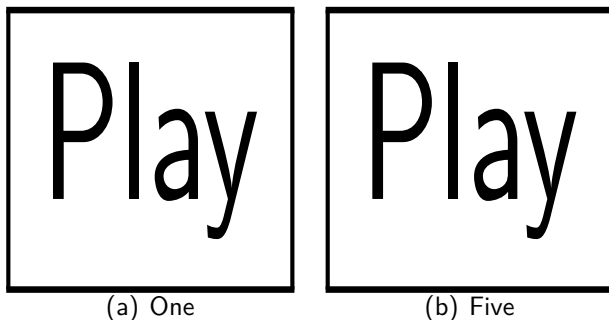
Figure : Recordings of someone saying "one" and "five."[1]

---

[1]S. Lafon, Y. Keller, and R. R. Coifman. Data Fusion and Multi-Cue Data Matching by Diffusion Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28, pages 1784-1797, 2006.

Lips
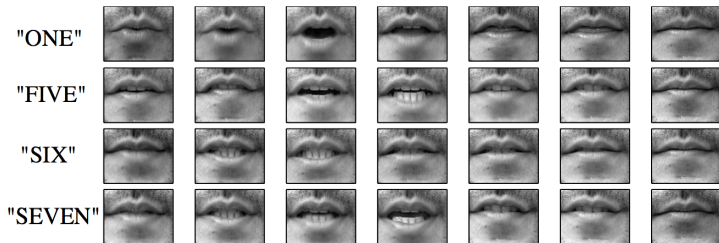
Figure : Typical frames for the words "one", "five", "six", "seven."[2]

[2]S. Lafon, Y. Keller, and R. R. Coifman. Data Fusion and Multi-Cue Data Matching by Diffusion Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28, pages 1784-1797, 2006.

The lips manifold

Figure : These manifold coordinates essentially capture two parameters: one controlling the opening of the mouth and the other measuring the portion of teeth that are visible.[3]

[3]S. Lafon, Y. Keller, and R. R. Coifman. Data Fusion and Multi-Cue Data Matching by Diffusion Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28, pages 1784-1797, 2006.
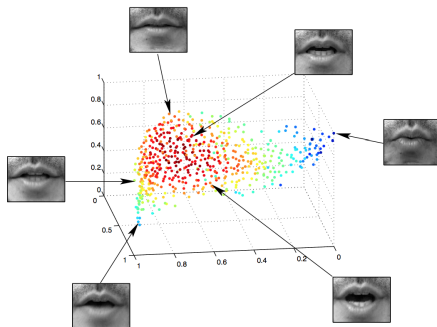
Diffusion maps overview
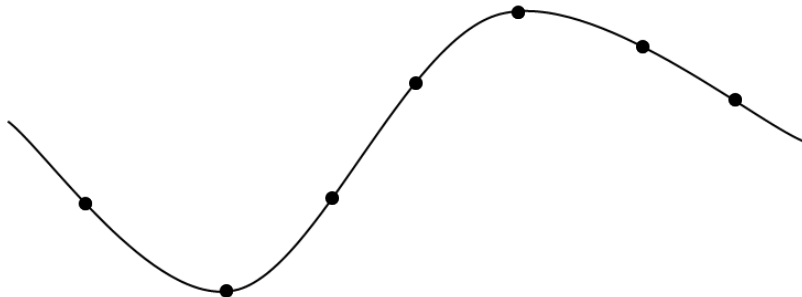
Gaussian kernel:

Diffusion maps overview

Gaussian kernel:

Diffusion maps overview
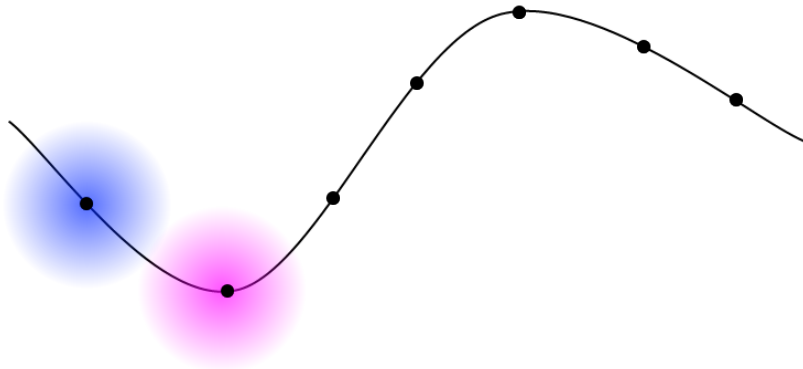
Gaussian kernel:

Diffusion maps overview
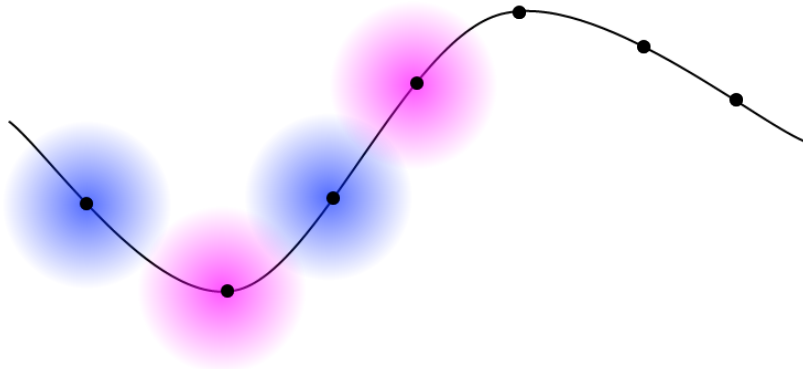
Gaussian kernel:

Diffusion maps overview
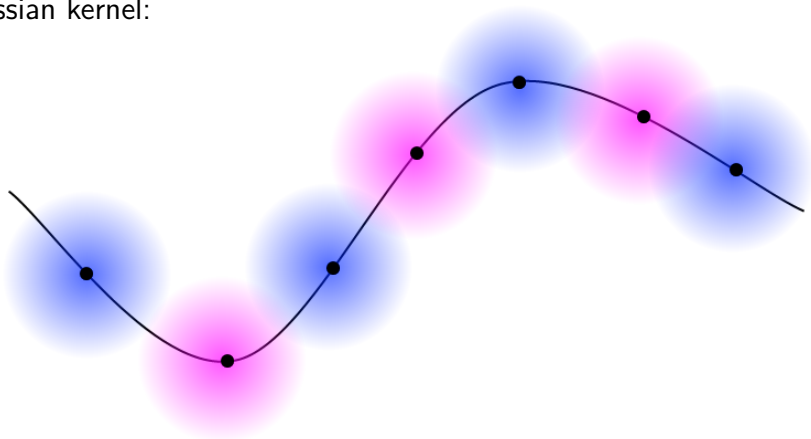
Gaussian kernel:

Diffusion maps overview
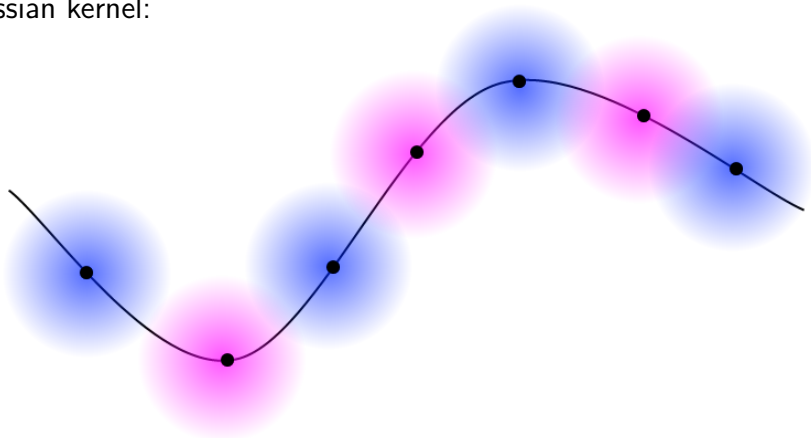
Gaussian kernel:

- Normalization $\implies$ diffusion kernel
- Spectral analysis $\implies$ map from $\mathcal{M} \subseteq \mathbb{R}^m$ to $\mathbb{R}^{\delta \ll m}$

Diffusion maps[4] overview

Diffusion process & affinities

- Gaussian kernel:
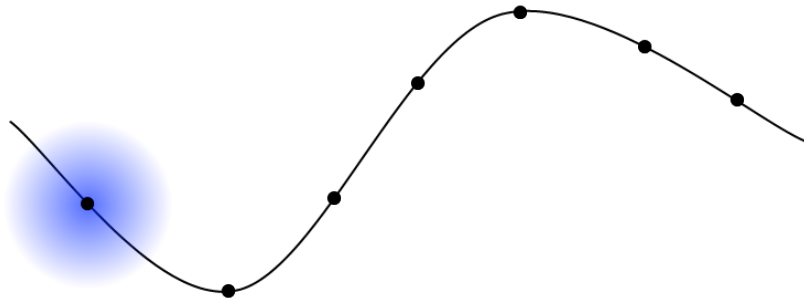 $$k(x, y) \triangleq e^{-\frac{\|x-y\|}{\varepsilon}}$$
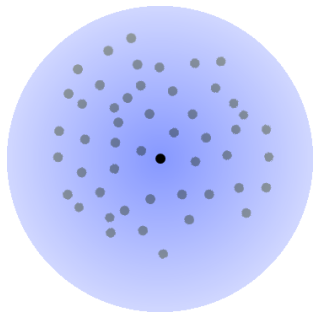
- Degrees: $q(x) \triangleq \sum k(x, y)$
- Transition probabilities:

$$p(x, y) \triangleq \frac{k(x, y)}{q(x)}$$

- Diffusion affinities:

$$a(x, y) \triangleq \frac{k(x, y)}{\sqrt{q(x)}\sqrt{q(y)}}$$
$$= q^{1/2}(x) p(x, y) q^{-1/2}(y)$$

[4]R.R. Coifman and S. Lafon. "Diffusion Maps". In: *Applied and Computational Harmonic Analysis* 21.1 (2006), pp. 5–30.

Spectrum (eigenvalues) of the diffusion affinity A and its powers
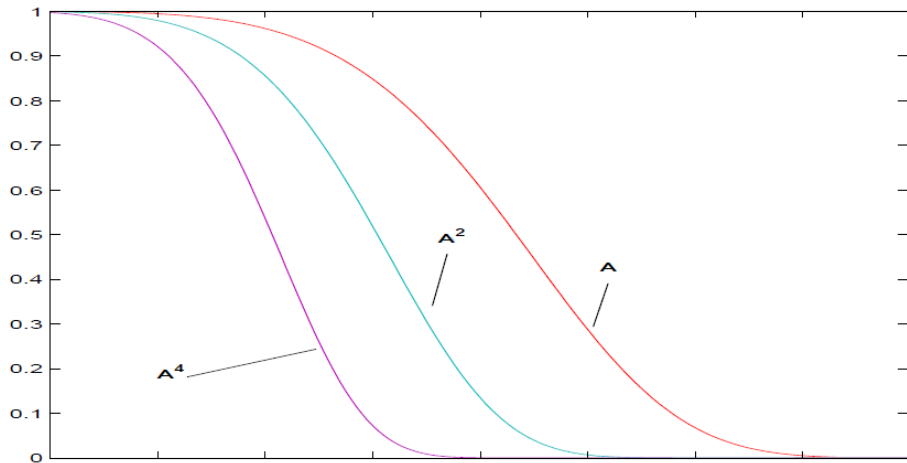
Spectrum (eigenvalues) of the diffusion affinity A and its powers

Diffusion maps overview
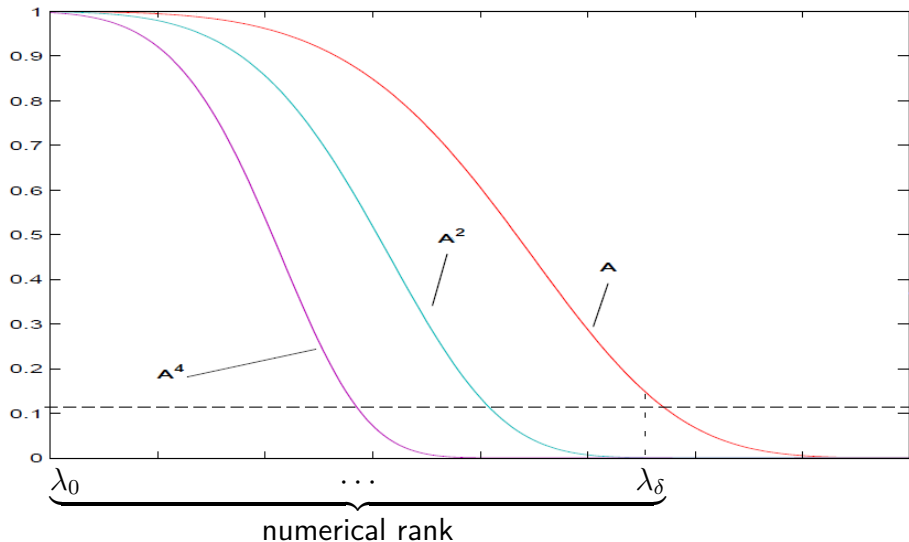
Spectral embedding

Diffusion maps overview
Spectral embedding

$$1 = \boxed{\lambda_0} \; \geq \; \boxed{\lambda_1} \; \geq \; \boxed{\lambda_2} \geq \; \cdots \; \geq \; \boxed{\lambda_\delta} > 0$$

$$n \left\{ \rule{0pt}{60pt} \right. \quad \boxed{\phi_0} \qquad \boxed{\phi_1} \qquad \boxed{\phi_2} \qquad \cdots \qquad \boxed{\phi_\delta}$$

Diffusion maps overview
Spectral embedding

$$1 = \boxed{\lambda_0} \;\geq\; \boxed{\lambda_1} \;\geq\; \boxed{\lambda_2} \geq\; \cdots \; \geq\; \boxed{\lambda_\delta} > 0$$
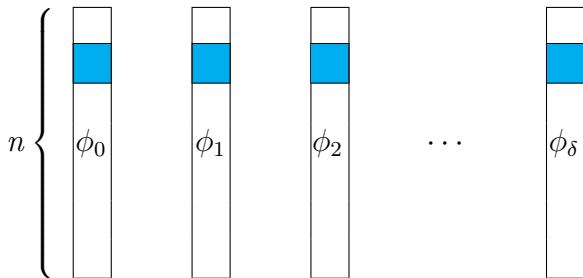
$$x \mapsto \Phi(x) \triangleq \left[\lambda_0 \phi_0(x)\,,\; \lambda_1 \phi_1(x)\,,\; \lambda_2 \phi_2(x)\,,\; \ldots\,,\; \lambda_\delta \phi_\delta(x) \right]^T$$

Diffusion maps overview

Embedded distances
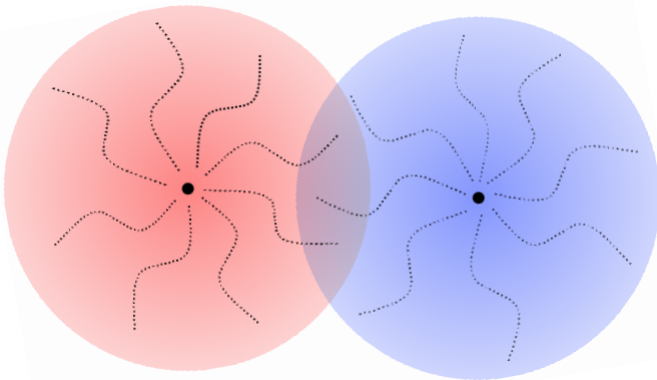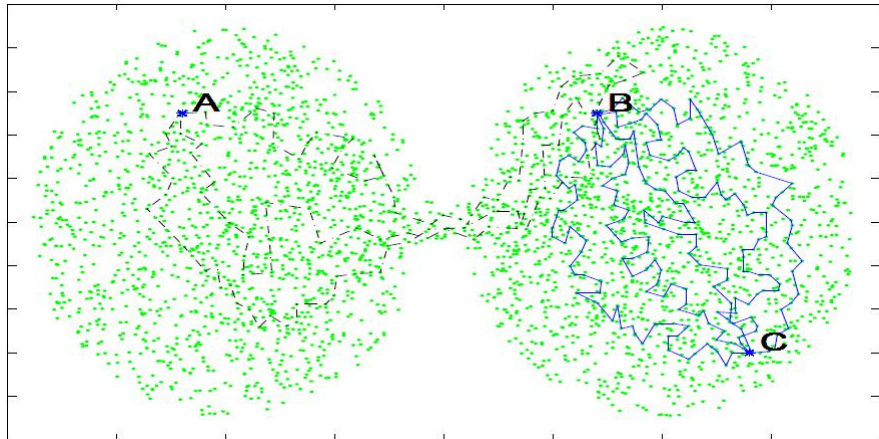
Overview

Notation

- $(\mathcal{M}, g)$ is a smooth, compact n-dimensional Riemannian manifold.
- Δ is the Laplace-Beltrami operator on $\mathcal{M}$.
- $L^2(\mathcal{M})$ is the space of functions $f : \mathcal{M} \to \mathbb{R}$ such that

$$\int_{\mathcal{M}} |f(x)|^2 \, dx < \infty.$$

- ℓ^2 is the space of real valued sequences $\{a_i\}_{i \geq 1}$ such that

$$\sum_{i=1}^{\infty} |a_i|^2 < \infty.$$

The heat equation

Definition

For a function $u(x, t)$ with domain and range $u : \mathcal{M} \times [0, T] \to \mathbb{R}$, the *heat equation* with initial solution $f(x)$ is

$$L(u) \triangleq \frac{\partial u}{\partial t} - \Delta_x u = 0$$

$$u(x, 0) = f(x)$$

- The fundamental solution of the heat equation is the heat kernel $K(t, x, y)$, which satisfies:
  - $L(K) = 0$ in (x, t).
  - $\lim_{t \to 0} K(t, x, y) = \delta_y(x)$.
- The solution $u(x, t)$ is given by:

$$u(x, t) = (e^{-t\Delta} f)(x, t) \triangleq \int_{\mathcal{M}} K(t, x, y) f(y) \, dy.$$

Diffusion distance

Definition

The *diffusion distance* at time t is defined as:

$$D_t(x, y)^2 \triangleq \int_{\mathcal{M}} (K(t, x, z) - K(t, y, z))^2 \, dz.$$
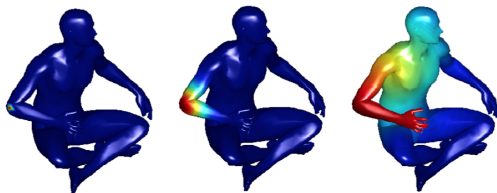
Figure : Heat kernel represents evolution of temperature.[5]

_____

[5]H. Abdallah. Processus de Diffusion sur un Flot de Variétés Riemanniennes. PhD thesis, L'Universite de Grenoble, 2010.

Spectral properties of the Laplacian

- The Laplacian Δ has a discrete sequence of eigenvalues

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \cdots \nearrow +\infty.$$

- Let $\{\psi_i\}_{i \geq 0} \subset L^2(\mathcal{M})$ be a set of orthonormal eigenfunctions of Δ.

- One can write the heat kernel as:

$$K(t, x, y) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \psi_i(x) \psi_i(y).$$

Embedding $\mathcal{M}$ into ℓ^2

Definition

For any $t > 0$, define the map $\Psi_t : \mathcal{M} \to \ell^2$ as:

$$\Psi_t(x) \triangleq \sqrt{2}(4\pi)^{\frac{n}{2}} t^{\frac{n+2}{4}} \left\{ e^{\frac{-\lambda_i t}{2}} \psi_i(x) \right\}_{i \geq 1}$$

Theorem (Bérard, Besson, and Gallot, 1994[6])

1. For all $t > 0$, the map Ψ_t is an embedding of $\mathcal{M}$ into ℓ^2.
2. The pulled-back metric Ψ_t^* is asymptotic to the metric g of $\mathcal{M}$ when $t \to 0$.
3. Ψ_t^* (up to the normalization factor) is the diffusion distance.

---

[6]P. Bérard, G. Besson, and S. Gallot. Embedding Riemannian manifolds by their heat kernel. *Geometric and Functional Analysis*, volume 4, number 4, pages 373-398, 1994.

Overview

Matthew Hirn and Guy Wolf   Diffusion Based Manifold Learning   October 23, 2013   22 / 88

Data, kernels, and graphs

- Our data set is X along with a measure μ that represents the distribution of the points.
- We assume that X lies on a low dimensional manifold $\mathcal{M}$.
- $k : X \times X \to \mathbb{R}$ is a symmetric, positive kernel.
- $k(x, y)$ measures the similarity between two points $x, y \in X$. The larger $k(x, y)$, the more similar.
- In high dimensions for points lying on a nonlinear manifold, we can only accurately measure local similarities, so k should behave accordingly. For example:

$$k(x, y) = k_\varepsilon(x, y) = e^{-\|x-y\|^2/\varepsilon}.$$

- X and k define a weighted graph $G = (X, k)$.

Starfish data

Figure : Data points sampled non-uniformly from a starfish shaped manifold.

Starfish weighted graph
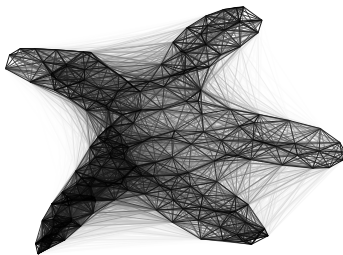
Figure : Graph with Gaussian weighted edges.

Markov chain

- To approximate heat diffusion on the underlying manifold, we construct a homogeneous Markov chain (random walk) on G.
- Define $d : X \to \mathbb{R}$ as:

$$d(x) \triangleq \int\limits_X k(x, y) \, d\mu(y).$$

- Define the transition kernel $p : X \times X \to \mathbb{R}$ of the random walk as:

$$p(x, y) \triangleq \frac{k(x, y)}{d(x)}.$$

- If G is connected (which we will assume), then the Markov chain has a unique stationary distribution $\pi : X \to \mathbb{R}$ given by:

$$\pi(x) = \frac{d(x)}{\int_X d(y) \, d\mu(y)}.$$

Markov chain

- Define the corresponding diffusion operator $P : L^2(X) \to L^2(X)$ as:

$$(Pf)(x) \triangleq \int_X p(x, y) f(y) \, d\mu(y).$$

- Powers P^t of the diffusion operator run the random walk forward in time:

$$(P^t f)(x) = \int_X p_t(x, y) f(y) \, d\mu(y),$$

where $p_t(x, y)$ is the probability of transition from x to y in t steps.

- Running the random walk forward will integrate the local geometry and reveal the relevant geometric structures of X at different scales.

Markov chain on Gaussian data
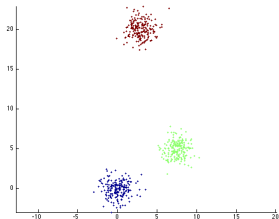
(a) Data consisting of 3 Gaussian clouds

(b) Powers of P

Diffusion distance

Definition

The *diffusion distance* at time t is defined as:

$$D_t(x, y)^2 \triangleq \int\limits_X \left(p_t(x, z) - p_t(y, z) \right)^2 \frac{d\mu(z)}{\pi(z)}.$$

Diffusion maps

- The operator P has a discrete sequence of eigenvalues

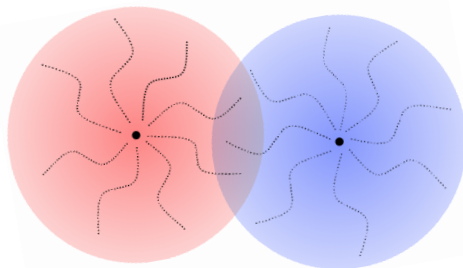$$1 = \lambda_0 > |\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \cdots \searrow 0.$$

- Let $\{\psi_i\}_{i \geq 0} \subset L^2(X)$ be a set of eigenfunctions of P.

Definition

For any $t > 0$, define the *diffusion map* $\Psi_t : X \to \ell^2$ as:

$$\Psi_t(x) \triangleq \{\lambda_i^t \psi_i(x)\}_{i \geq 1}$$

Theorem (Coifman and Lafon, 2006[7])

The diffusion map preserves the diffusion distance:

$$D_t(x, y) = \|\Psi_t(x) - \Psi_t(y)\|_{\ell^2}$$

[7]R. R. Coifman and S. Lafon. Diffusion Maps. *Applied and Computational Harmonic Analysis*, volume 21, pages 5-30, 2006.

Vehicle trajectory
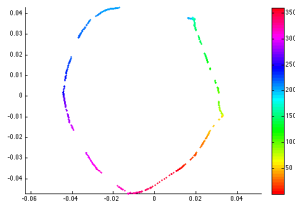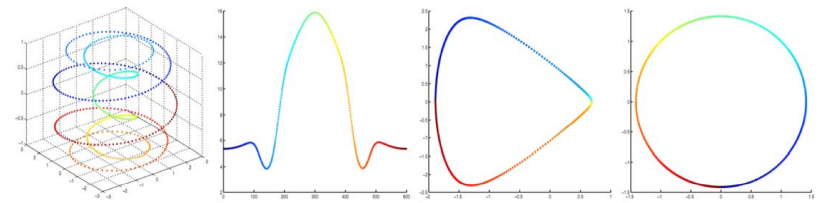

(c) BTR-60   (d) BTR-60 driving in a circle

Figure : Diffusion map of the BTR video, restricted to the first two eigenvectors. The color corresponds to the angle on the circle at which the BTR is located.

Statistics vs geometry

From the left[8]:

1. A closed curve sampled non-uniformly.
2. Density of the samples.
3. Diffusion embedding of the curve in which the sampling skews the geometry.
4. A modified diffusion embedding of the curve in which the statistics have been separated from the geometry.

[8]R. R. Coifman and S. Lafon. Diffusion Maps. *Applied and Computational Harmonic Analysis*, volume 21, pages 5-30, 2006.
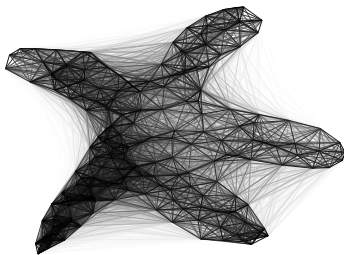
Figure : Graph with Gaussian weighted edges.

Modified Markov chain

- Let $k_\varepsilon(x, y) = e^{-\|x-y\|^2/\varepsilon}$, $x, y \in X = \mathcal{M}$.
- Approximate the density $q(x)dx = d\mu(x)$ via

$$q_\varepsilon(x) \triangleq \int\limits_X k_\varepsilon(x, y) q(y) \, dy.$$

- Form a new kernel

$$k_\varepsilon^{(\alpha)}(x, y) \triangleq \frac{k_\varepsilon(x, y)}{q_\varepsilon(x)^\alpha q_\varepsilon(y)^\alpha}, \qquad \alpha \in [0, 1]$$

- Normalize $k_\varepsilon^{(\alpha)}$ to be a Markov chain by defining

$$d_\varepsilon^{(\alpha)}(x) \triangleq \int\limits_X k_\varepsilon^{(\alpha)}(x, y) q(y) \, dy,$$

and setting

$$p_\varepsilon^{(\alpha)}(x, y) \triangleq \frac{k_\varepsilon^{(\alpha)}(x, y)}{d_\varepsilon^{(\alpha)}(x)}.$$

Infinitesimal generator

- Define the diffusion operator $P_\varepsilon^{(\alpha)} : L^2(X) \to L^2(X)$,

$$(P_\varepsilon^{(\alpha)} f)(x) \triangleq \int\limits_X p_\varepsilon^{(\alpha)}(x, y) f(y) q(y) \, dy.$$

Theorem (Coifman and Lafon, 2006[9])

Let

$$L_\varepsilon^{(\alpha)} = \frac{I - P_\varepsilon^{(\alpha)}}{\varepsilon},$$

be the infinitesimal generator of the Markov chain. Then,

$$\lim_{\varepsilon \to 0} L_\varepsilon^{(\alpha)} f = \frac{\Delta(f q^{1-\alpha})}{q^{1-\alpha}} - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}} f.$$

[9]R. R. Coifman and S. Lafon. Diffusion Maps. *Applied and Computational Harmonic Analysis*, volume 21, pages 5-30, 2006.

Corollaries

Two interesting cases:

- $\alpha = 0$. This was the original construction. In this case:

$$\lim_{\varepsilon \to 0} L_\varepsilon^{(0)} f = \frac{\Delta(fq)}{q} - \frac{\Delta(q)}{q} f,$$

  and the density q fully influences the infinitesimal generator.

- $\alpha = 1$. In this case the density q is removed and we recover the Laplacian:

$$\lim_{\varepsilon \to 0} L_\varepsilon^{(1)} f = \Delta f.$$

This in turn implies that $P_\varepsilon^{(1)}$ approximates the heat kernel integral operator:

$$\lim_{\varepsilon \to 0} \left(P_\varepsilon^{(1)} \right)^{\frac{t}{\varepsilon}} = e^{-t\Delta}.$$
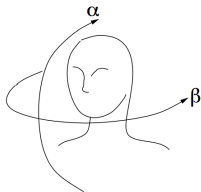
Head rotation data

Figure : Each subject essentially moved their head along the two angles α and β. There was almost no tilting of the head.
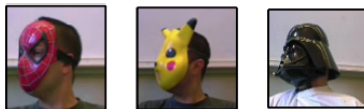

Figure : The three subjects each with their head rotated at a particular orientation.[10]

Head rotation data embeddings
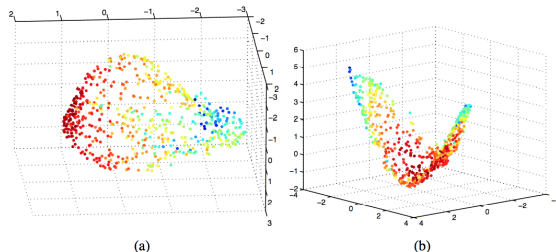

(a)   (b)

Figure : The embeddings of the YELLOW (a) and BLACK (b) sets in three diffusion coordinates without the density renormalization. The color encodes the density of the points. These embedded sets have very different shapes, and their alignment is impossible.[11]

[11]S. Lafon, Y. Keller, and R. R. Coifman. Data Fusion and Multi-Cue Data Matching by Diffusion Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28, pages 1784-1797, 2006.

Head rotation data embeddings

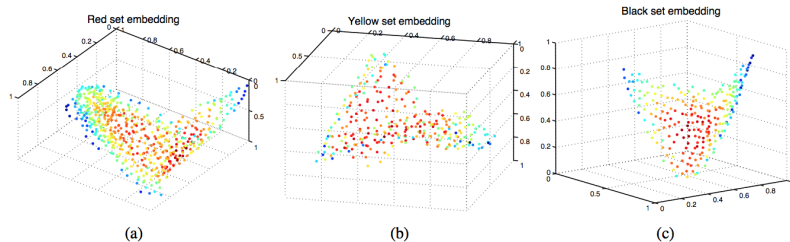(a)   (b)   (c)

Figure : The embedding of each set in the first three diffusion coordinates with the density renormalization. All three sets share this butterfly shaped embedding.[12]

---

[12]S. Lafon, Y. Keller, and R. R. Coifman. Data Fusion and Multi-Cue Data Matching by Diffusion Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28, pages 1784-1797, 2006.
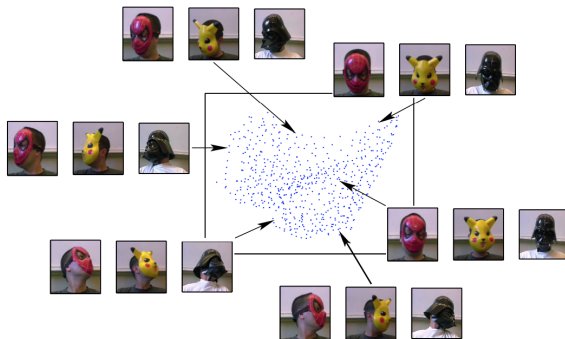
Head rotation data embeddings

Figure : The embeddings of the YELLOW, RED, and BLACK sets after alignment.[13]

---

[13]S. Lafon, Y. Keller, and R. R. Coifman. Data Fusion and Multi-Cue Data Matching by Diffusion Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 28, pages 1784-1797, 2006.

Discrete approximations

- Sample $\{x_1, \ldots, x_m\} \subset X$ according to the density q.
- We have the following discrete approximations of the previous functions and operators:

$$\overline{q}_\varepsilon(x_i) \triangleq \sum_{j=1}^m k_\varepsilon(x_i, x_j)$$

$$\overline{k}_\varepsilon^{(\alpha)}(x_i, x_j) \triangleq \frac{k_\varepsilon(x_i, x_j)}{q_\varepsilon(x_i)^\alpha q_\varepsilon(x_j)^\alpha}$$

$$\overline{d}_\varepsilon^{(\alpha)}(x_i) \triangleq \sum_{j=1}^m \overline{k}_\varepsilon^{(\alpha)}(x_i, x_j)$$

$$\overline{p}_\varepsilon^{(\alpha)}(x_i, x_j) \triangleq \frac{\overline{k}_\varepsilon^{(\alpha)}(x_i, x_j)}{\overline{d}_\varepsilon^{(\alpha)}(x_i)}$$

$$(\overline{P}_\varepsilon^{(\alpha)} f)(x_i) \triangleq \sum_{j=1}^m \overline{p}_\varepsilon^{(\alpha)}(x_i, x_j) f(x_j)$$

Convergence

Theorem (Singer, 2006[14])

The convergence rate from the discrete to the continuous Markov chain is, with high probability,
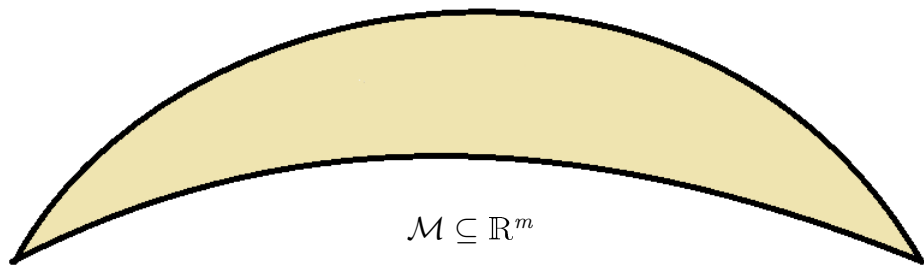
$$|(\overline{P}_\varepsilon^{(\alpha)} f)(x_i) - (P_\varepsilon^{(\alpha)} f)(x_i)| = O(m^{-\frac{1}{2}} \varepsilon^{-\frac{n}{4} - \frac{1}{2}}).$$

---

[14]A. Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, volume 21, pages 128-134, 2006.
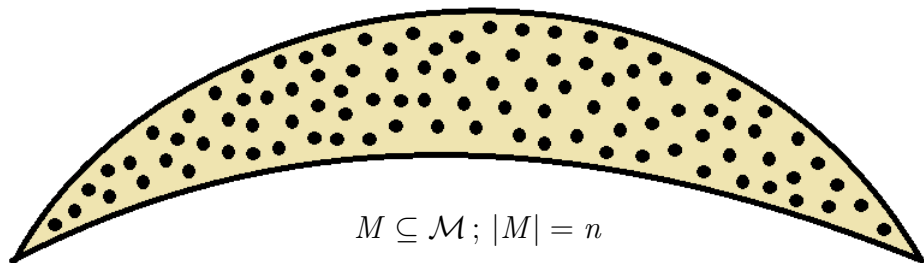
Overview

Patch-to-Tensor Embedding[15]

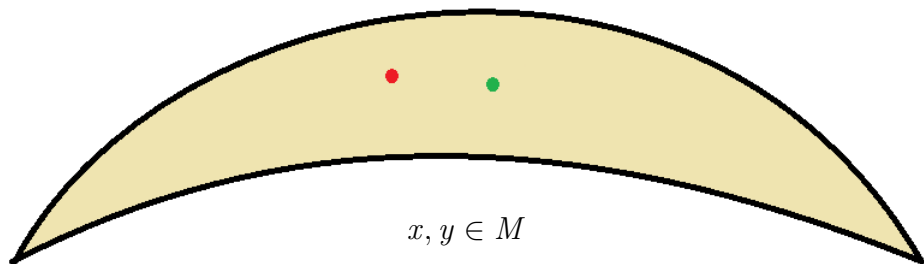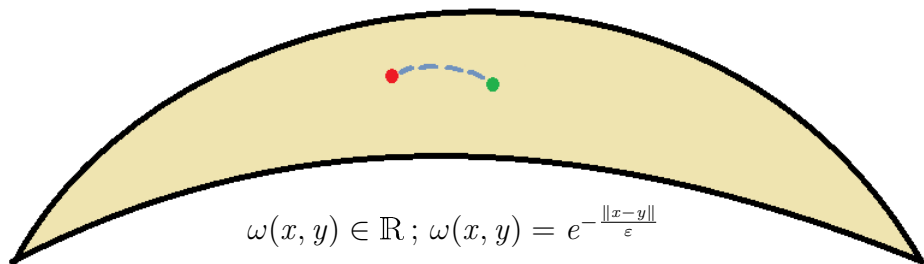$$\mathcal{M} \subseteq \mathbb{R}^m$$

[15]M. Salhov, GW, and A. Averbuch. "Patch-to-Tensor Embedding". In: *Applied and Computational Harmonic Analysis* 33.2 (2012), pp. 182 –203.

Patch-to-Tensor Embedding[15]

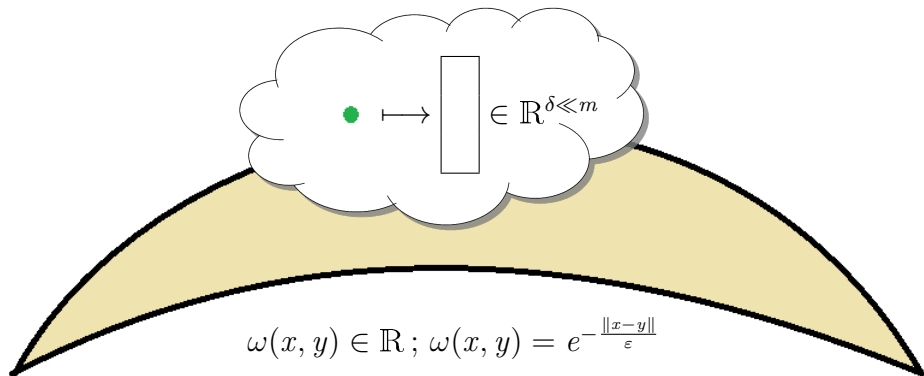$$M \subseteq \mathcal{M} \, ; \, |M| = n$$

[15]M. Salhov, GW, and A. Averbuch. "Patch-to-Tensor Embedding". In: *Applied and Computational Harmonic Analysis* 33.2 (2012), pp. 182 –203.

Patch-to-Tensor Embedding[15]
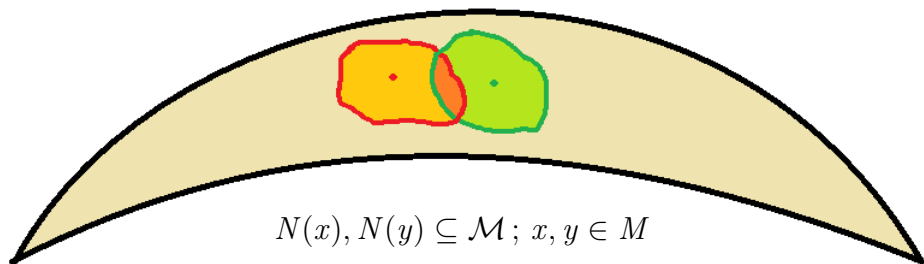

$$x, y \in M$$

[15]M. Salhov, GW, and A. Averbuch. "Patch-to-Tensor Embedding". In: *Applied and Computational Harmonic Analysis* 33.2 (2012), pp. 182 –203.

Patch-to-Tensor Embedding[15]

$$\omega(x, y) \in \mathbb{R} \ ; \ \omega(x, y) = e^{-\frac{\|x-y\|}{\varepsilon}}$$

[15]M. Salhov, GW, and A. Averbuch. "Patch-to-Tensor Embedding". In: *Applied and Computational Harmonic Analysis* 33.2 (2012), pp. 182 –203.

$$\omega(x,y) \in \mathbb{R} \; ; \; \omega(x,y) = e^{-\frac{\|x-y\|}{\varepsilon}}$$
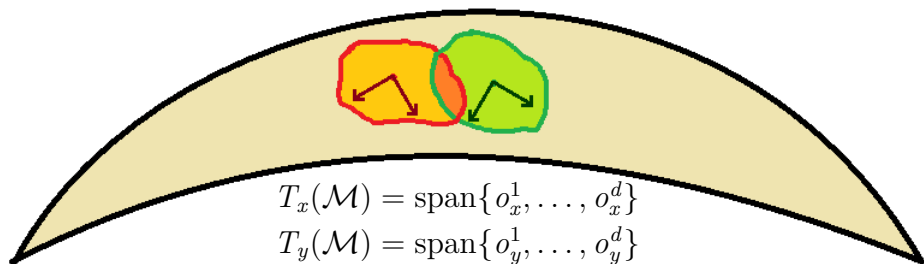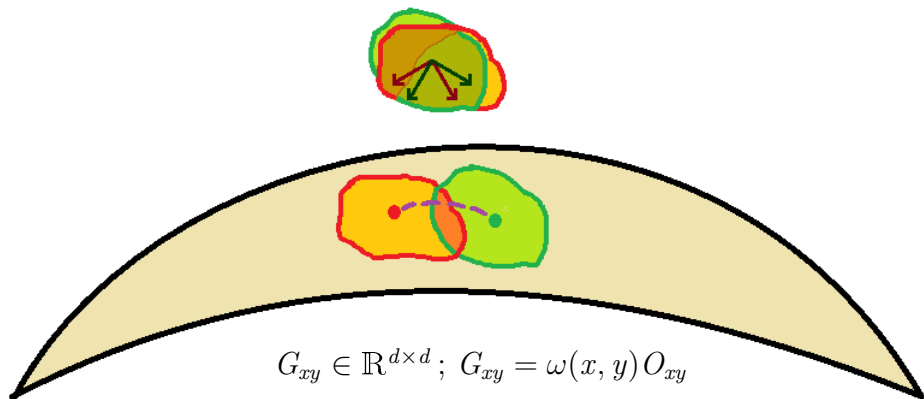
[15]M. Salhov, GW, and A. Averbuch. "Patch-to-Tensor Embedding". In: *Applied and Computational Harmonic Analysis* 33.2 (2012), pp. 182 –203.

Patch-to-Tensor Embedding[15]

$$N(x), N(y) \subseteq \mathcal{M} \, ; \, x, y \in M$$

[15]M. Salhov, GW, and A. Averbuch. "Patch-to-Tensor Embedding". In: *Applied and Computational Harmonic Analysis* 33.2 (2012), pp. 182 –203.
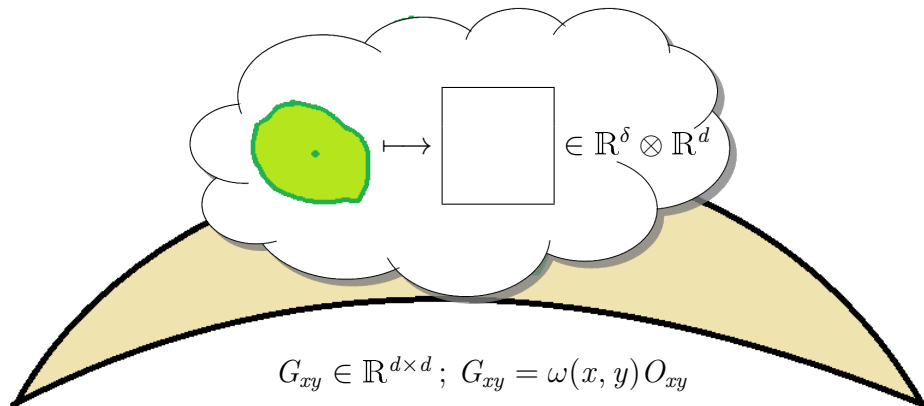
$$T_x(\mathcal{M}) = \mathrm{span}\{o_x^1, \ldots, o_x^d\}$$
$$T_y(\mathcal{M}) = \mathrm{span}\{o_y^1, \ldots, o_y^d\}$$

[15]M. Salhov, GW, and A. Averbuch. "Patch-to-Tensor Embedding". In: *Applied and Computational Harmonic Analysis* 33.2 (2012), pp. 182 –203.

Patch-to-Tensor Embedding[15]

$$G_{xy} \in \mathbb{R}^{d \times d}; \; G_{xy} = \omega(x, y) O_{xy}$$

[15]M. Salhov, GW, and A. Averbuch. "Patch-to-Tensor Embedding". In: Applied and Computational Harmonic Analysis 33.2 (2012), pp. 182 –203.

Patch-to-Tensor Embedding[15]

$$G_{xy} \in \mathbb{R}^{d \times d}; \; G_{xy} = \omega(x, y) O_{xy}$$

[15]M. Salhov, GW, and A. Averbuch. "Patch-to-Tensor Embedding". In: *Applied and Computational Harmonic Analysis* 33.2 (2012), pp. 182 –203.

Super-kernel

Vector field representation

A vector field can be represented in two equivalent ways:

Super-kernel
Definition (block matrix)

A super-kernel G is a block matrix / operator on vector-fields:

Super-kernel
Operator on vector-fields

The super-kernel as an operator on vector-fields:

$$\vec{v}(x) \in T_x(\mathcal{M}) \overset{G}{\mapsto} \sum G_{xy}\vec{v}(y) = G\vec{v}(x) \in T_x(\mathcal{M})$$

Diffusion super-kernel: $\omega(x, y) = $ the diffusion affinity $a(x, y)$

Linear-Projection Diffusion[16]

Diffusion step

[16]GW and A. Averbuch. "Linear-Projection Diffusion on Smooth Euclidean Submanifolds". In: *Applied and Computational Harmonic Analysis* 34.1 (2013), pp. 1 –14.

Linear-Projection Diffusion[16]

Diffusion step

[16] GW and A. Averbuch. "Linear-Projection Diffusion on Smooth Euclidean Submanifolds". In: *Applied and Computational Harmonic Analysis* 34.1 (2013), pp. 1 –14.
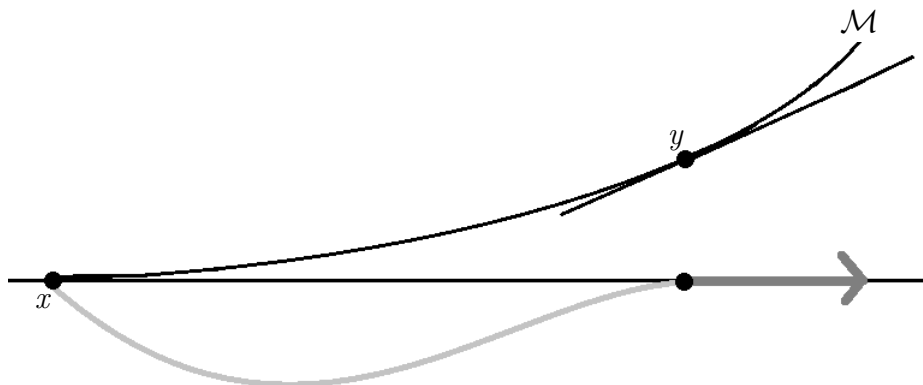
Linear-Projection Diffusion[16]

Diffusion step

[16]GW and A. Averbuch. "Linear-Projection Diffusion on Smooth Euclidean Submanifolds". In: *Applied and Computational Harmonic Analysis* 34.1 (2013), pp. 1 –14.
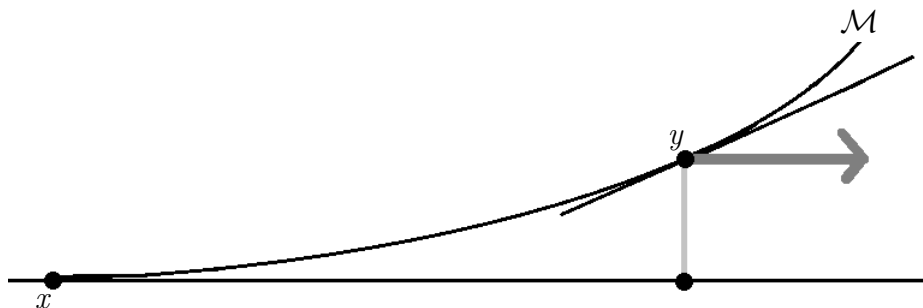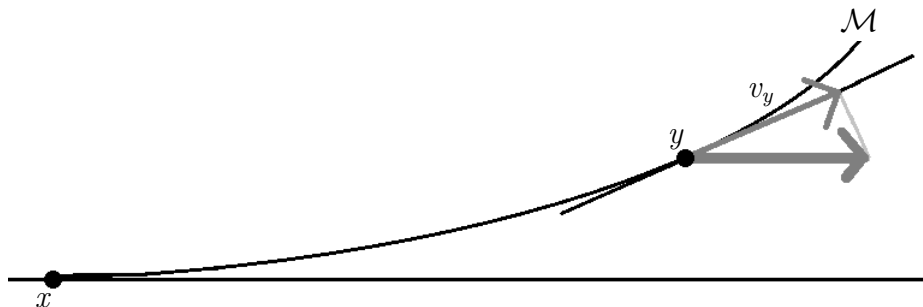
Linear-Projection Diffusion[16]

Diffusion step

[16] GW and A. Averbuch. "Linear-Projection Diffusion on Smooth Euclidean Submanifolds". In: *Applied and Computational Harmonic Analysis* 34.1 (2013), pp. 1 –14.

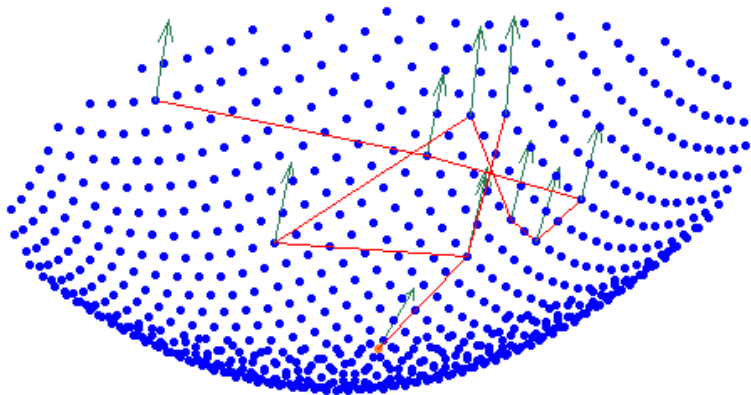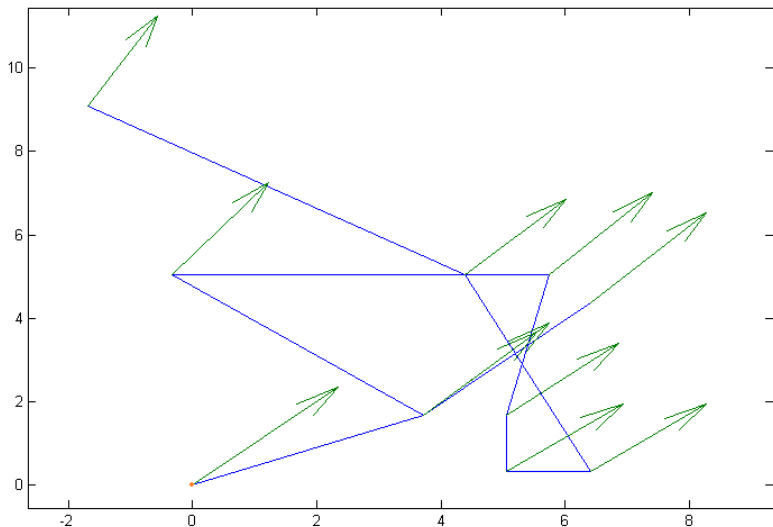**Infinitesimal generator** - a **partial derivative operator** that incorporates significant properties of a stochastic process

- Isotropic
- Anisotropic
-

Infinitesimal Generator:

$$\mathcal{L} = \lim_{\varepsilon \to 0} \frac{I - P_\varepsilon}{\varepsilon}$$

Linear-Projection Diffusion

Infinitesimal generator

Infinitesimal generator - a **partial derivative operator** that incorporates significant properties of a stochastic process

- Isotropic Diffusion Maps: Graph Laplacian
- Anisotropic Diffusion Maps: Laplace-Beltrami
- Vector Diffusion

Heat Equation:

$$\frac{\partial u}{\partial t} = \alpha \nabla^2 u$$

Infinitesimal generator - a **partial derivative operator** that incorporates significant properties of a stochastic process

- Isotropic Diffusion Maps: **Graph Laplacian**
- Anisotropic Diffusion Maps: **Laplace**-**Beltrami**
- Vector Diffusion Maps: **Connection Laplacian**

Infinitesimal generator - a **partial derivative operator** that incorporates significant properties of a stochastic process

- Isotropic Diffusion Maps: **Graph Laplacian**
- Anisotropic Diffusion Maps: **Laplace**-**Beltrami**
- Vector Diffusion Maps: **Connection Laplacian**

- Linear-Projection Diffusion:
  Vector-Laplacian of tangent projections

Linear-Projection (LP) super-kernel consists of the blocks:

$$\underbrace{\boxed{G_{xy}}}_{d \times d} = \underbrace{\boxed{\omega}}_{(x,y)} \quad \overbrace{\boxed{O_x^T}}^{d \times m} \quad \left.\boxed{O_y}\right\} p \times m$$

LPD super-kernel: $\omega(x, y) =$ the diffusion affinity $a(x, y)$

Linear-Projection Diffusion (LPD) Super-Kernel

Ambient cooridnate independence

Proposition

The Linear-Projection super-kernel is independent of the coordinates of the ambient space.

Proof.

Consider a change of basis $O_x \mapsto BO_x$, where B is an orthogonal matrix:

$$
\begin{aligned}
G_{xy} = \omega(x, y)(BO_x)^T(BO_y) &= \omega(x, y)O_x^T B^T B O_y \\
&= \omega(x, y)O_x^T O_y \qquad x, y \in \mathcal{M}.
\end{aligned}
$$

Therefore, the blocks G_{xy} remain the same in every basis of the ambient space. $\square$

Linear-Projection Diffusion (LPD) Super-Kernel
Spectral properties

Ω = Scalar affinity kernel $\qquad$ G = Linear-Projection super-kernel

- Ω is positive (semi) definite $\Rightarrow$ G is positive (semi) definite

- Ω has operator norm $\|\Omega\| \Rightarrow G$ as operator norm $\|G\| \le \|\Omega\|$
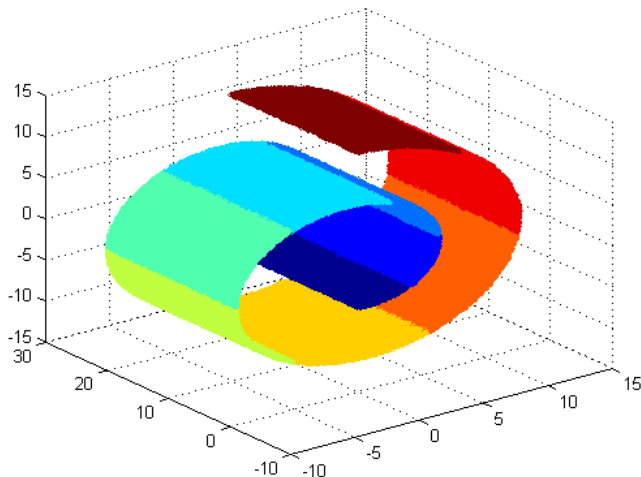
Spectrum of the LPD super-kernel

$$\Omega = A$$

$$\Downarrow$$

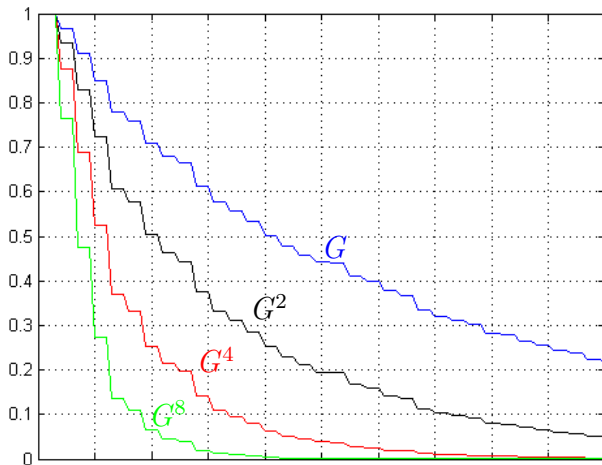$$\|A\| = 1 \ge \lambda_0 \ge \lambda_1 \ge \lambda_2 \ge \ldots > 0$$

LPD-based Patch-to-Tensor Embedding
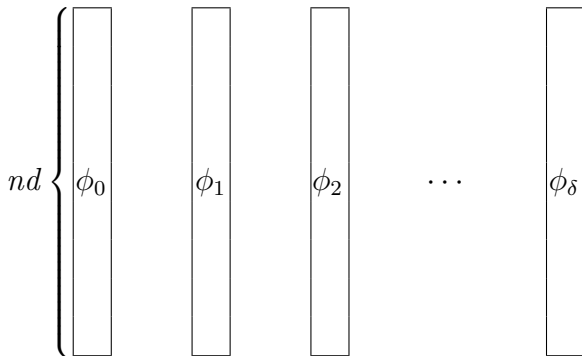
Spectral embedding

Spectrum (eigenvalues) of the LPD super-kernel G and its powers

LPD-based Patch-to-Tensor Embedding
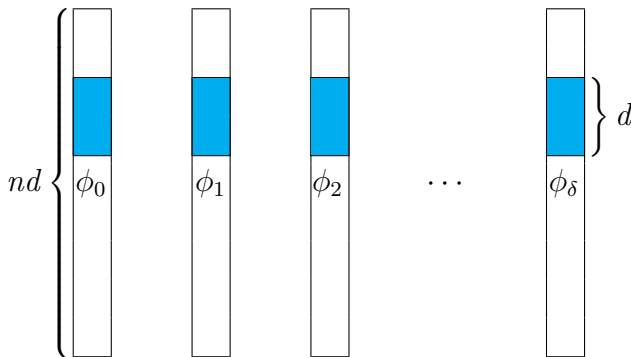Spectral embedding

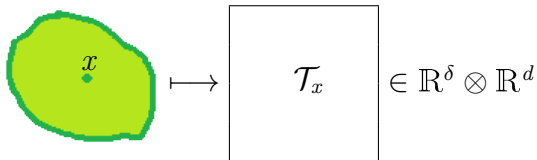$$1 \geq \boxed{\lambda_0} \geq \boxed{\lambda_1} \geq \boxed{\lambda_2} \geq \cdots \geq \boxed{\lambda_\delta} > 0$$

$$nd \left\{ \quad \phi_0 \qquad \phi_1 \qquad \phi_2 \qquad \cdots \qquad \phi_\delta \right.$$

LPD-based Patch-to-Tensor Embedding

Spectral embedding

LPD-based Patch-to-Tensor Embedding

Embedded tensors

Embedded tensors

$$
\mathcal{T}_x = \left.\left(\begin{array}{c} \boxed{\lambda_1^t} \cdot \boxed{[\; -\; \vec{\phi}_1(x)\; -\;]} \\ \\ \vdots \\ \\ \boxed{\lambda_\delta^t} \cdot \boxed{[\; -\; \vec{\phi}_\delta(x)\; -\;]} \end{array}\right)\right\} \delta \times d
$$

$$
x \longmapsto \boxed{\mathcal{T}_x} \in \mathbb{R}^\delta \otimes \mathbb{R}^d
$$

LPD-based Patch-to-Tensor Embedding

Extended diffusion distances

PTE with the diffusion super-kernel

$$\overbrace{\|\mathcal{T}_x - \mathcal{T}_y\|_F^2}^{\text{embedded distances}} = \overbrace{\sum_{z \in M} \|a(x,z)O_{xz} - a(y,z)O_{yz}\|_F^2}^{\text{extended diffusion distances}}$$

PTE with the LPD super-kernel

$$\overbrace{\|\mathcal{T}_x - \mathcal{T}_y\|_F^2}^{\text{embedded distances}} = \overbrace{\sum_{z \in M} \|a(x,z)O_x^T O_z - a(y,z)O_y^T O_z\|_F^2}^{\text{extended diffusion distances}}$$

$$= \underbrace{\sum_{z \in M} \sum_{j=1}^{d} \|a(x,z)O_x^T o_z^j - a(y,z)O_y^T o_z^j\|^2}_{\text{LPD distances}}$$
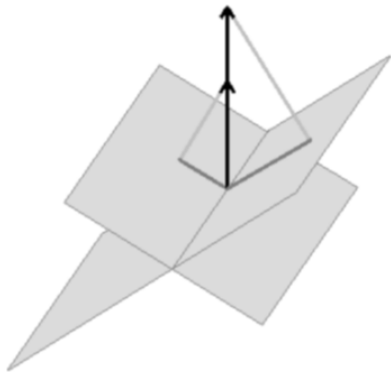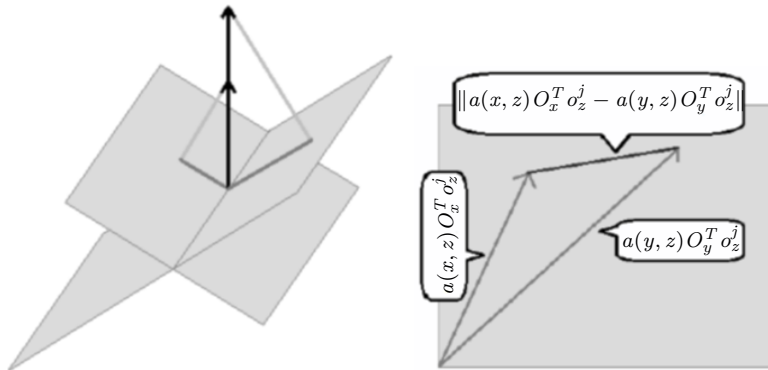
LPD-based Patch-to-Tensor Embedding

LPD distances

$$\|\mathcal{T}_x - \mathcal{T}_y\|_F^2 = \sum_{z \in M} \sum_{j=1}^{d} \|a(x,z) O_x^T o_z^j - a(y,z) O_y^T o_z^j\|^2$$
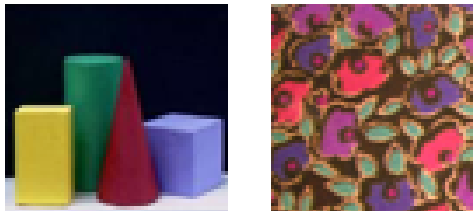
LPD-based Patch-to-Tensor Embedding
LPD distances

$$\|\mathcal{T}_x - \mathcal{T}_y\|_F^2 = \sum_{z \in M} \sum_{j=1}^{d} \|a(x,z) O_x^T o_z^j - a(y,z) O_y^T o_z^j\|^2$$

$$\|\mathcal{T}_x - \mathcal{T}_y\|_F^2 = \sum_{z \in M} \sum_{j=1}^{d} \|a(x,z) O_x^T o_z^j - a(y,z) O_y^T o_z^j\|^2$$

$H \times W$ RGB image $\longrightarrow (H \cdot W) \times 5$ dataset

- Patches of pixels $\xrightarrow{\text{PTE \& LPD}}$ embedded tensors
- Cluster the patches/tensors based on the embedded distances
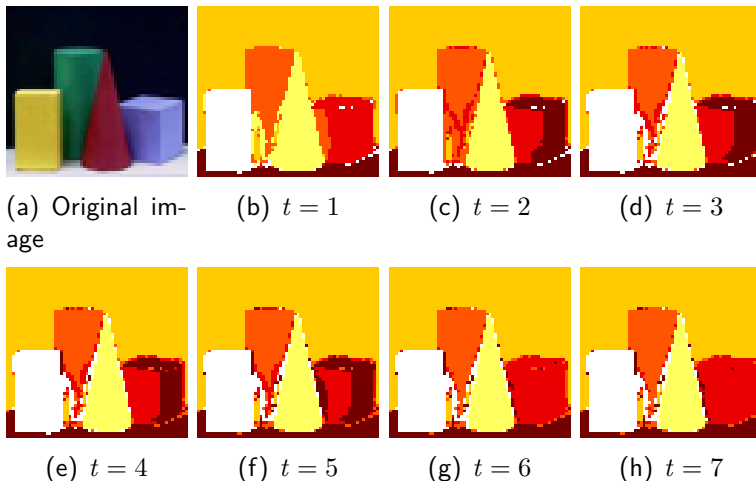
Image segmentation
Achieved results

(a) Original image

(b) $t = 1$

(c) $t = 2$

(d) $t = 3$

(e) $t = 4$

(f) $t = 5$

(g) $t = 6$

(h) $t = 7$

Figure : The PTE segmentation results shown at several diffusion times t.

Image segmentation

Additional results

(a) Original image

(b) $t = 1$

(c) $t = 2$

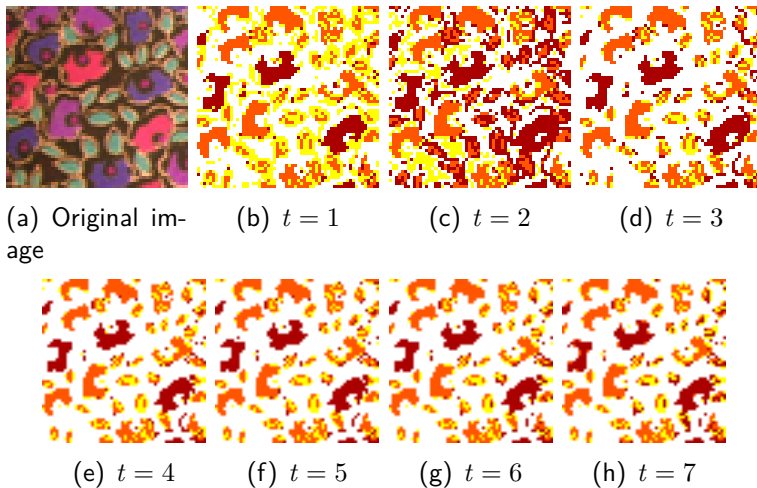(d) $t = 3$

(e) $t = 4$

(f) $t = 5$

(g) $t = 6$

(h) $t = 7$

Figure : The PTE segmentation results shown at several diffusion times t.

Overview

Change detection

- Two images, different time acquisition.
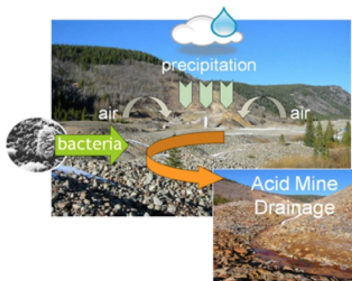- Where did the changes occur?
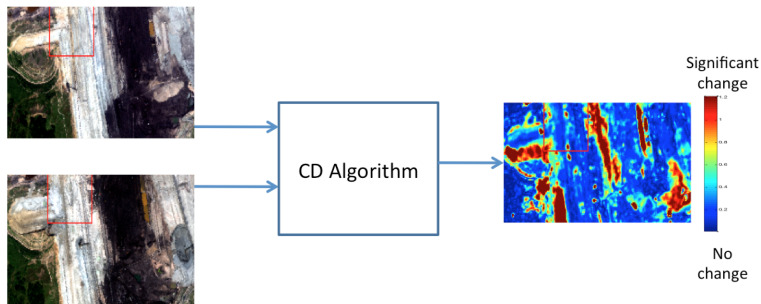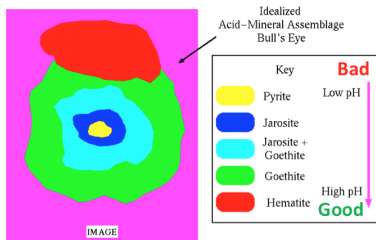
- Water acidity causes:
  - Vegetation stress and loss of biodiversity
  - Better transfer of heavy metals and other toxic components

[17]M. A. Sares, P. L. Hauff, D. C. Peters, D. W. Coulter, D. A. Bird, F. B. Henderson III, and E. C. Prosh. Characterizing Sources of Acid Rock Drainage and Resulting Water Quality Impacts Using Hyperspectral Remote Sensing - Examples from the Upper Arkansas River Basin, Colorado. *Advanced integration of geospatial technologies in mining reclamation*, December 7-9, 2004, Atlanta, Georgia, USA.
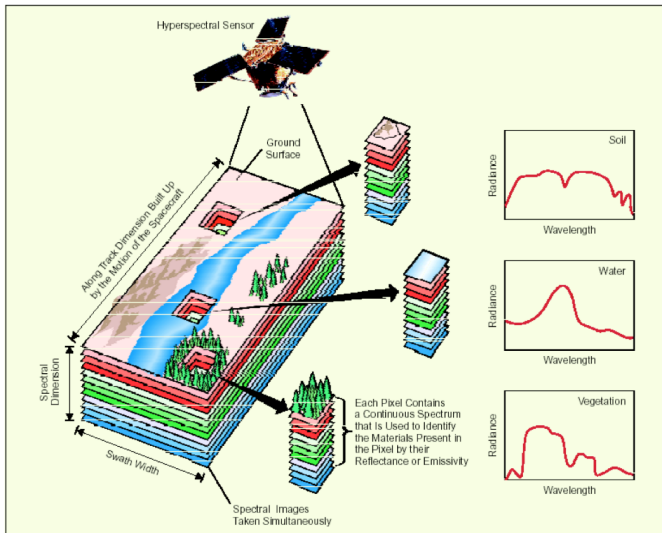
- Mining controlling authorities
  - Field chemical measurements
  - Area is too big
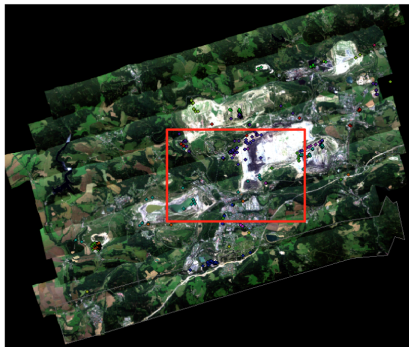
High cost of monitoring

6

Solution: Imaging spectrometer

Sokolov coal mine area mosaic images

HyMap 2009

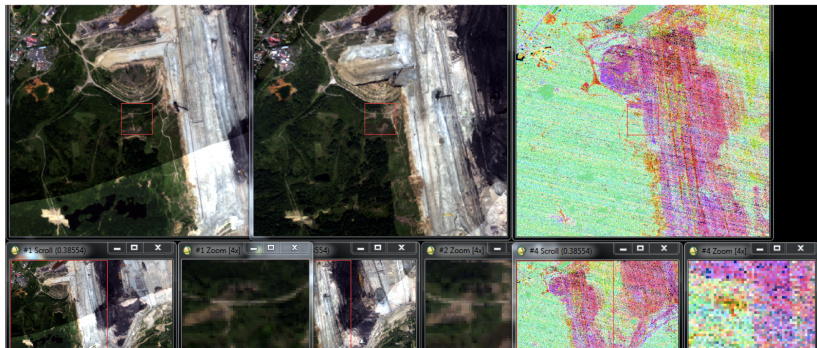HyMap 2010

Sokolov coal mine

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

Sample spectral signatures
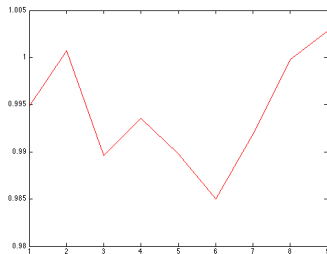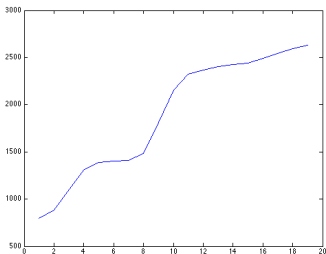
(a) Spectral signature of a vegetation pixel, 2009, during the day, visible range.

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

No direct comparison

Direct comparison

High dimensional data at time 1

High dimensional data at time 2

Compare neighborhoods

Compare neighborhoods

High dimensional data at time 1

High dimensional data at time 2
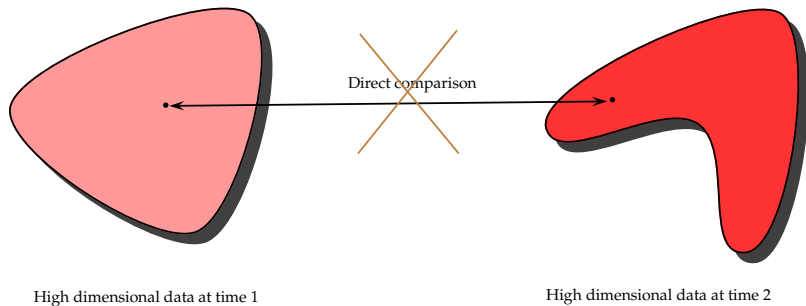
Data

- We are given several data sets $X_\tau = (X, d_\tau, \mu)$ indexed by $\tau \in \mathcal{I}$. Each data set is a metric measure space in which the metric depends on the parameter τ.
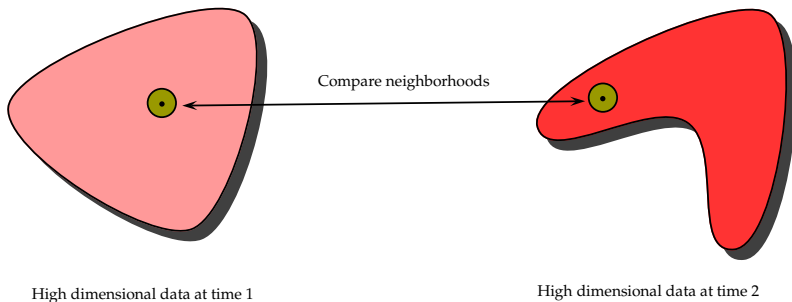
- We think of $\mathcal{I}$ as the set of times or parameters for which we view/measure/observe (X, μ).

- The changes in (X, μ) across $\mathcal{I}$ are encoded in the metrics $\{d_\tau\}_{\tau \in \mathcal{I}}$.

Markov chains and diffusion distance

- For each $\tau \in \mathcal{I}$ we construct a homogenous Markov chain $P^{(\tau)}$ along with the corresponding transition kernel $p^{(\tau)}$.
- Let $\pi^{(\tau)}$ be the stationary distribution of $P^{(\tau)}$.
- Let $\{\lambda_i^{(\tau)}\}_{i\geq 0}$ denote the eigenvalues of $P^{(\tau)}$, and let $\{\psi_i^{(\tau)}\}_{i\geq 0}$ be a set of corresponding eigenfunctions.

Definition

The *diffusion* distance at time t between $x_\tau \in X_\tau$ and $y_\sigma \in X_\sigma$ is defined as:

$$D_t(x_\tau, y_\sigma)^2 \triangleq \int\limits_X \left(\frac{p_t^{(\tau)}(x,z)}{\sqrt{\pi^{(\tau)}(z)}} - \frac{p_t^{(\sigma)}(y,z)}{\sqrt{\pi^{(\sigma)}(z)}} \right)^2 \, d\mu(z).$$

Diffusion maps

- For each $\tau \in \mathcal{I}$ we have the diffusion map

$$\Psi_t^{(\tau)}(x) \triangleq \left\{ \left(\lambda_i^{(\tau)} \right)^t \psi_i^{(\tau)}(x) \right\}_{i \geq 1}.$$
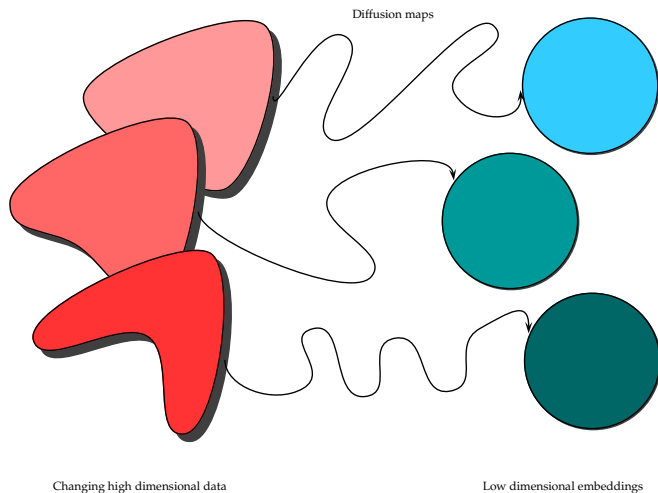
- The standard diffusion maps theory tells us that

$$D_t(x_\tau, y_\tau) = \left\| \Psi_t^{(\tau)}(x) - \Psi_t^{(\tau)}(y) \right\|_{\ell^2}.$$

- On the other hand, if $\tau \neq \sigma$, then

$$D_t(x_\tau, y_\sigma) \neq \left\| \Psi_t^{(\tau)}(x) - \Psi_t^{(\sigma)}(y) \right\|_{\ell^2}.$$

Multiple low dimensional embeddings

Diffusion maps

Changing high dimensional data

Low dimensional embeddings

Symmetric diffusion operators

- Symmetrize $p^{(\tau)}$ by conjugating with the squareroot of the density function $d^{(\tau)}$,

$$a^{(\tau)}(x, y) \triangleq \sqrt{d^{(\tau)}(x)} \cdot p^{(\tau)}(x, y) \cdot \frac{1}{\sqrt{d^{(\tau)}(y)}}.$$

- Define the corresponding integral operator as:

$$(A^{(\tau)}f)(x) \triangleq \int_X a^{(\tau)}(x, y)f(y) \, d\mu(y).$$

- The eigenvalues of $A^{(\tau)}$ and $P^{(\tau)}$ are the same. Let $\{\varphi_i^{(\tau)}\}_{i \geq 0}$ denote the eigenfunctions of $A^{(\tau)}$.

Aligning the embeddings

- Let $a = \{a_i\}_{i \geq 0} \in \ell^2$.
- For $\tau, \sigma \in \mathcal{I}$, define the map $O_{\tau \to \sigma} : \ell^2 \to \ell^2$ as:

$$O_{\tau \to \sigma}(a) \triangleq \left\{ \sum_{j \geq 0} a_j \cdot \langle \varphi_j^{(\tau)}, \varphi_i^{(\sigma)} \rangle \right\}_{i \geq 0}.$$
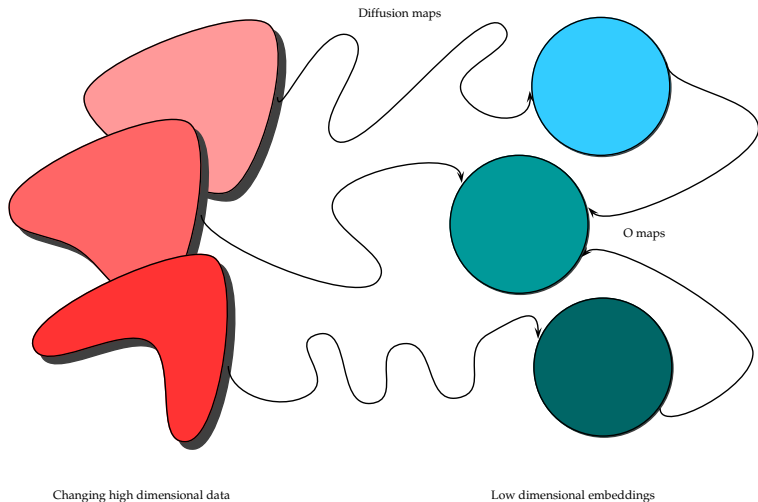
Theorem (Coifman and H., 2013[18])

Fix $\gamma \in \mathcal{I}$. Then, for any $\tau, \sigma \in \mathcal{I}$, $t > 0$, and $x_\tau \in X_\tau$, $y_\sigma \in X_\sigma$,

$$D_t(x_\tau, y_\sigma) = \left\| O_{\tau \to \gamma} \Psi_t^{(\tau)}(x) - O_{\sigma \to \gamma} \Psi_t^{(\sigma)}(y) \right\|_{\ell^2}.$$

---

[18]R. R. Coifman and M. H. Diffusion maps for changing data. *Applied and Computational Harmonic Analysis*, In press.

Single embedding

Diffusion maps

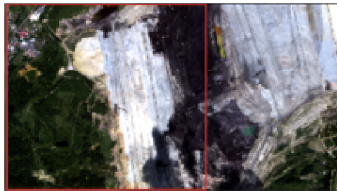O maps

Changing high dimensional data
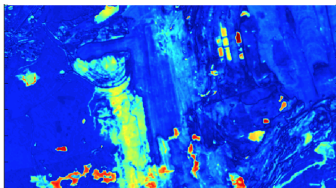
Low dimensional embeddings

Change detection using diffusion maps

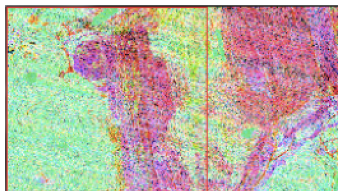(c) Sokolov Mine 2009 / Day / Visible / HyMap sensor

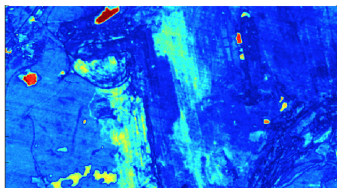(d) Sokolov Mine 2011 / Day / Visible / AHS sensor

(e) Change map using diffusion maps.
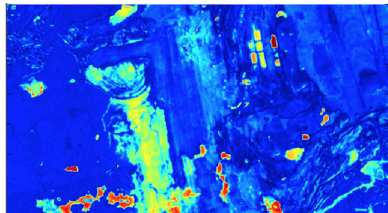
Change detection using diffusion maps

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

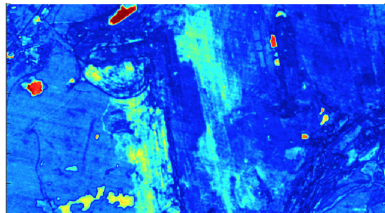(g) Sokolov Mine 2011 / Night / Thermal / AHS sensor

(h) Change map using diffusion maps

Comparison of change maps

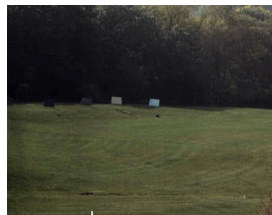(i) 2009 / Day / Visible / HyMap sensor vs. 2011 / Day / Visible / AHS sensor

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

(k) September

(l) October

(m) $D^{(t)}(x_{\mathsf{Sep}}, x_{\mathsf{Oct}})$ as $t \to \infty$

The Standard Map

- The standard map is an area preserving chaotic map from a square with side length 2π onto itself.
- It is a discrete difference equation defined as:

$$p_{i+1} = p_i + K \cdot \sin(\theta_i)$$
$$\theta_{i+1} = \theta_i + p_{i+1}$$

- $K = 0$: the map is linear, and only periodic and quasi-periodic orbits are possible.
- $K > 0$: the map is nonlinear, and certain initial conditions give chaotic dynamics. The larger K, the more chaotic.
- Used to model what can happen in certain plasma driven fusion experiments.

Ergodic quotient

- Let $x = (x_0, x_1, x_2, \ldots)$ be a discrete trajectory on a manifold $\mathcal{M}$ with $\dim(\mathcal{M}) = n$.

- Test the trajectory x against Fourier modes:

$$\phi_x(\omega) = \sum_{j=0}^{\infty} e^{2\pi i \omega \cdot x_j}, \quad \omega \in \mathbb{Z}^n.$$

- The distance between two trajectories x and y is then defined as:

$$d(x, y) = \sum_{\omega \in \mathbb{Z}^n} \frac{|\phi_x(\omega) - \phi_y(\omega)|^2}{(1 + |\omega|^2)^s},$$

where $s = (n+1)/2$.

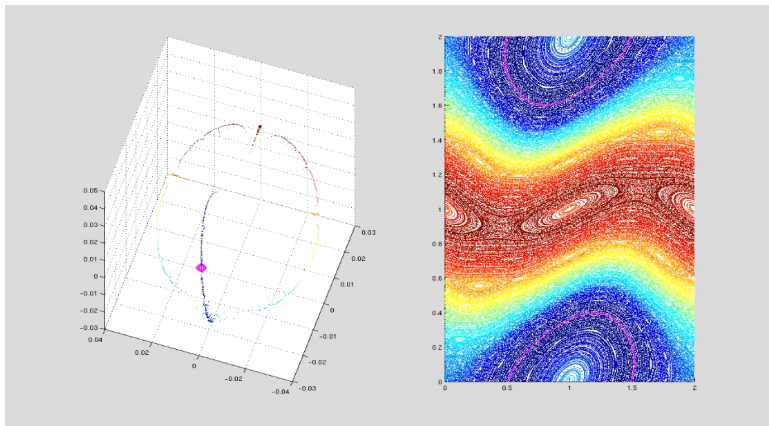# Diffusion Embedding of the Standard Map

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.

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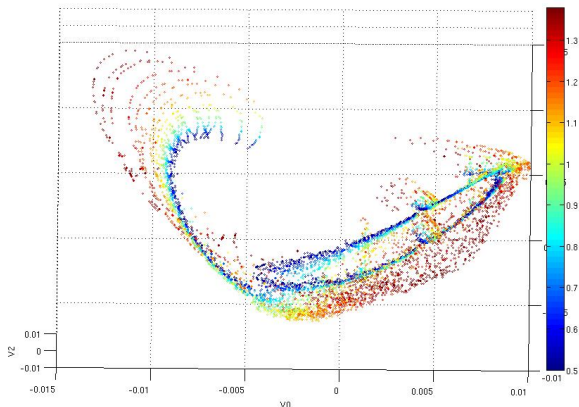
Common embedding

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).

The End