Spectral Graph Theory and Diffusion Wavelets

Marie Wild

Date/time: February/5th/2007, 16.00

Location: Seminar room 183/2, Favoritenstr. 9, 4th floor

ABSTRACT:

This talk is meant to provide an overview of the general framework of spectral graph techniques and their (existing and possible) applications for computer vision tasks. This includes the recent concept of diffusion wavelets, developed by R. R. Coifman, M. Maggioni et al., which allows wavelet analysis on graphs. The talk is planned in order to wake your interest on these topics and should serve as a basis for subsequent discussion.

DISCLAIMER:

As this talk is meant to serve as an introduction to the above techniques rather than a presentation of the authors own scientific work, all the **results (including pictures) are borrowed** from the sources and authors cited therein.

Spectral Graph Theory and Diffusion Wavelets

PRIP, TU Wien

February 5, 2007

Marie Wild

smart systems

AUSTRIAN RESEARCH CENTERS

Motivation of this talk

- Introduction on the abstract framework of spectral graph theory (and wavelets on graphs as a subtopic)

- Glance on existing and POSSIBLE applications in Computer Vision (Clustering, Segmentation, Tracking)

$\Longrightarrow$ Convince you that these issues are worth to be studied
$\Longrightarrow$ Looking for people interested in collaboration

Outline of this talk

1. Spectral Graph Theory - Global Analysis
   (a) Basics: Diffusion on a Graph
   (b) Relation to Random Walks
   (c) Relation to Fourier Analysis
   (d) Applications in Computer Vision
2. Spectral Graph Theory - Multiscale Analysis
   (a) Wavelets and Multiresolution Analysis on $\mathbb{R}$
   (b) Multiresolution Analysis on a Graph: Diffusion Wavelets
   (c) Examples
3. Conclusion and POSSIBLE FURTHER APPLICATIONS

1. Spectral Graph Theory- Global Analysis

FAN R.K. CHUNG: **SPECTRAL GRAPH THEORY**, CMBS-AMS, NO. 92, 1997

• Properties and structure of a graph from the spectrum (Eigenvalues) of a matrix describing local properties of a graph, e.g. connectedness, bottlenecks

• Random walk interpretation (global information from local information)

• Interpretation in terms of Fourier Analysis $\longrightarrow$ Extension to Wavelet Analysis possible (Information on multiple scales from local information)

R.R.COIFMAN, S. LAFON, A.B. LEE, M. MAGGIONI, B. NADLER, F. WARNER AND S.W. ZUCKER: **GEOMETRIC DIFFUSIONS AS A TOOL FOR HARMONIC ANALYSIS AND STRUCTURE DEFINITION OF DATA**, PROC. NAT. AC. OF SC., VOL 102(21), MAY 2005

R.R. COIFMAN, M. MAGGIONI: **DIFFUSION WAVELETS**, ACHA, VOL. 21(1), JULY 2006

(a) Basics: Diffusion on a graph

- $G = (V, E)$ undirected graph with weight matrix $W : V \times V \to \mathbb{R}^+$,
- $d_v = \sum_{u \in \mathcal{V}} W(u, v)$ degree of vertex v.
- (normalized) Laplacian on G:

$$L = \begin{cases} 1 - \frac{W(v,v)}{d_v} & \text{if } u = v \text{ and } d_v \neq 0, \\ -\frac{W(u,v)}{\sqrt{d_u d_v}} & \text{if } u, v \text{ adjacent}, \\ 0 & \text{otherwise}. \end{cases}$$

$L = I - D^{-1/2} W D^{-1/2}$, D diagonal matrix with entries d_v.

- Diffusion on G:

$$K = I - L.$$

L (or K) describes local similarity in the graph.
Global properties can be explored from its Eigenvalues.

Let n be the number of vertices in G.
L is symmetric and positive semidefinite, $\lambda_i \geq 0$ for all $i = 0, \ldots, n-1$.
Let $0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{n-1}$

- Example (Connectivity):

If G is connected, then $\lambda_1 > 0$. If $\lambda_i = 0$ and $\lambda_{i+1} \neq 0$, then G has exactly $i+1$ components.

(b) Random Walk Interpretation

$\sum_{v \in V} K(u,v) = 1 \Rightarrow$
K can be interpreted as transition matrix of a random walk on G,

$K^m(u,v)$ represents the probability of walking from u to v in m steps.

'Explore the graph by walking on it'

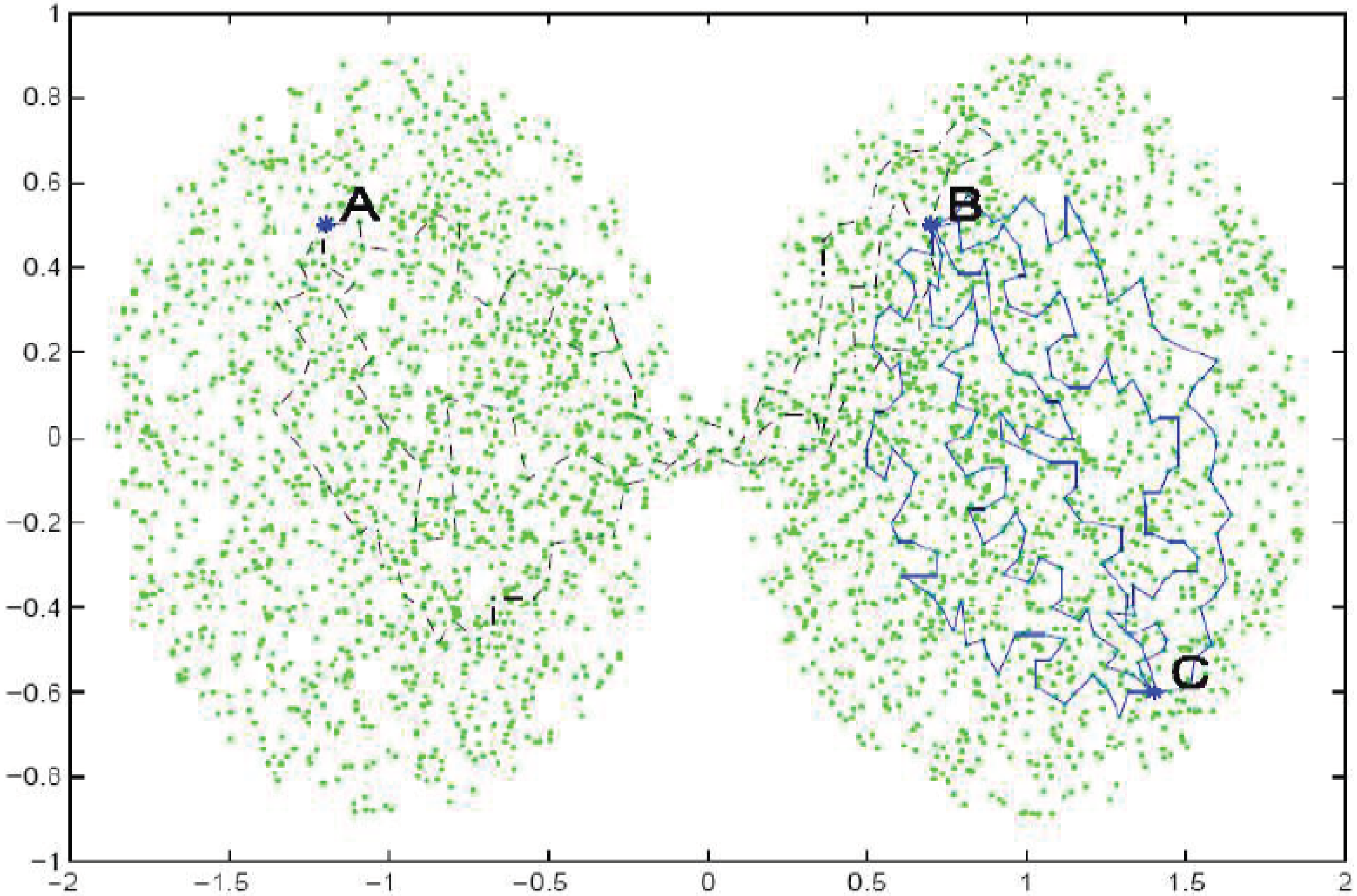
Properties of this random walk (stationary distribution, speed of convergence) can be explored by the spectrum.

- Diffusion distance, COIFMAN ET. AL.:

$$D^m(u,v) = \|K^m(u,\cdot) - K^m(v,\cdot)\|_2$$

Measures the strength of **all** paths between vertices $\simeq$ likelihood of getting from one vertex to another $\Longrightarrow$ **robustness** to noise

Diffusion vs. Geodesic Distance



$d_{geod.}(A,B) \sim d_{geod.}(C,B)$, however $d^{(t)}(A,B) >> d^{(t)}(C,B)$.

Picture courtesy of S. Lafon

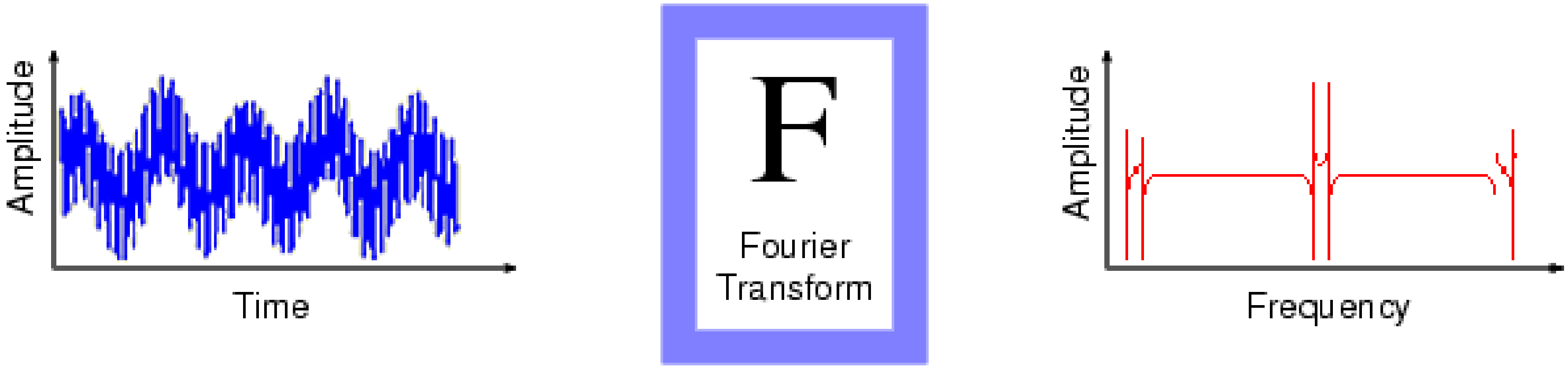
(b) Fourier Analytic Interpretation

- Fourier Transform in Engineering: Discrete Fourier Transform (DFT)

$$\text{Discrete-Time, finite signal } (x_n)_{0\le n\le N} \xrightarrow{\text{FFT}} (\widehat{x}_k)_{0\le k\le N}$$

$\widehat{x}_k = \sum_n x_n e^{\frac{2\pi i}{N}kn}$, Frequency representation of (x_n).

$x_n = \frac{1}{N}\sum_k \widehat{x}_k e^{-\frac{2\pi i}{N}kn}$

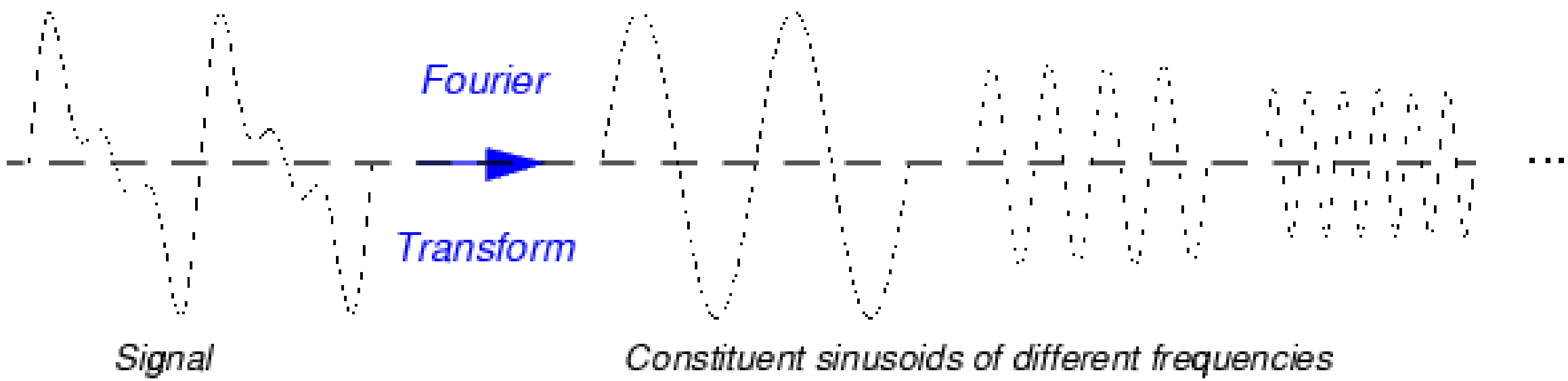


• Mathematical Formulation: DFT corresponds to discretization of Fourier Series

For $f \in L^2(\mathcal{T})$,

$$f(t) = \sum_{n \in \mathbb{Z}} \langle f, e^{2\pi itn} \rangle e^{2\pi itn},$$

$(e^{2\pi itn})_{n\in\mathbb{Z}}$ ONB in $L^2(\mathcal{T})$



The Fourier basis is known to diagonalize certain operators T
(among them Convolution operators, Laplace and Diffusion operators), furthermore

$T(e^{2\pi itn}) = \lambda_n(e^{2\pi itn}) \quad \Rightarrow e^{2\pi itn}$ Eigenvectors of T.

Leads to natural generalization of Fourier analysis on a graph G:

Go the other way round: define the eigenvectors ϕ_i (sorted by decreasing eigenvalues) of the diffusion operator K as generalized Fourier basis functions $\Rightarrow$

for $f \in L^2(G)$

$$f = \sum_{i \in I} \langle f, \phi_i \rangle \phi_i,$$

the larger i, the more oscillating ϕ_i and λ_i^{-1} measures the frequency of ϕ_i

$\Rightarrow$ Fourier analysis on a graph

This can be used to organize **local information** into **global parametrization**:

• back to the Diffusion distance :

$$D^m(u,v) = \|K^m(u,\cdot) - K^m(v,\cdot)\|_2 = \left(\sum_i \lambda_i^{2m} |\phi_i(u) - \phi_i(v)|^2\right)^{1/2}$$

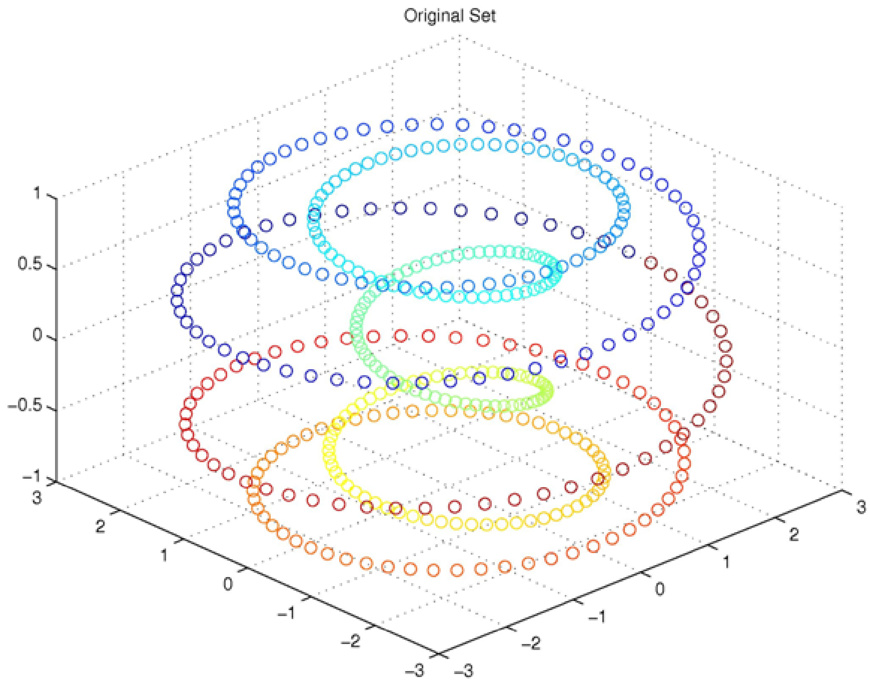
$\Rightarrow$ Diffusion map: The (nonlinear) embedding

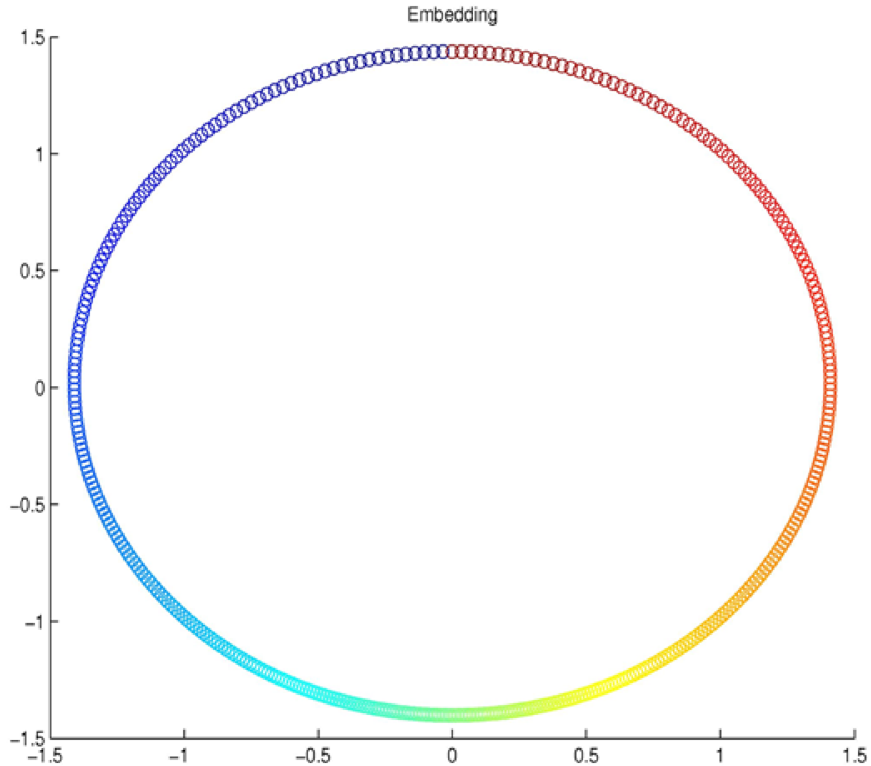
$$x \in G \hookrightarrow X(x) = \{\lambda_i \phi_i\} \in l^2$$

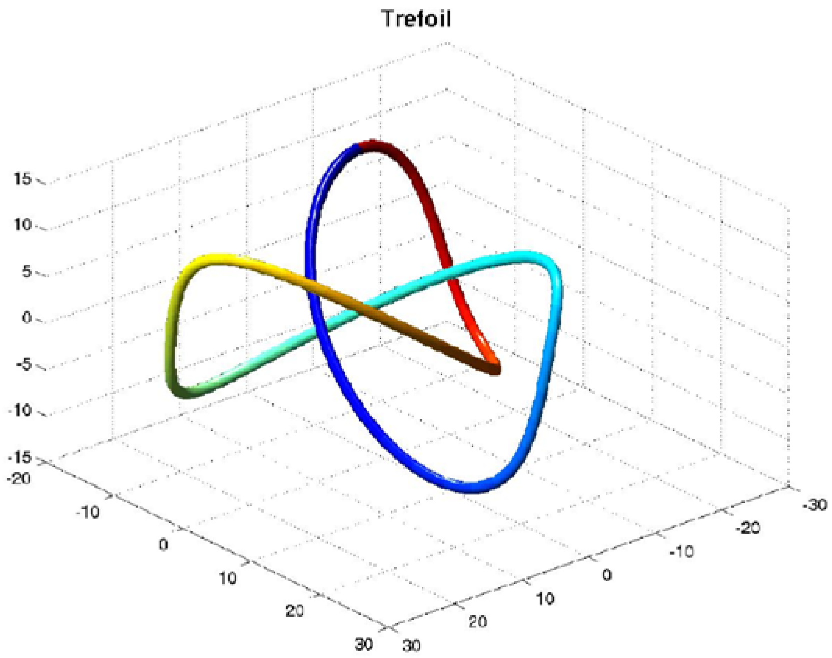
maps diffusion distance to euclidean distance

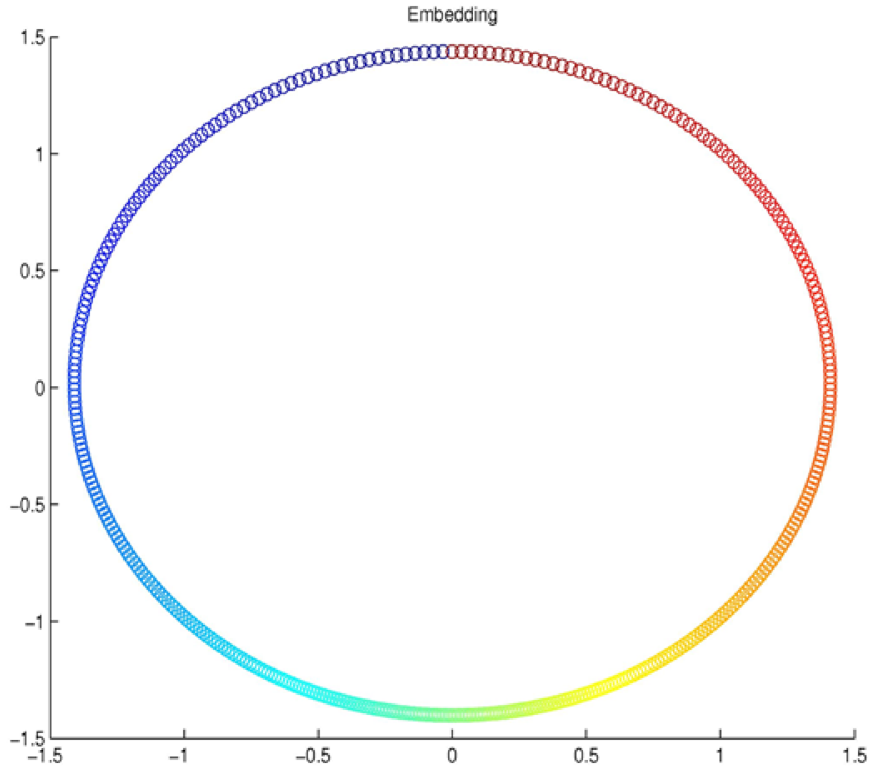
frequency grows with $i \Rightarrow$ diffusion distance can be aprroximately calculated by a truncated series, using only a few eigenvectors

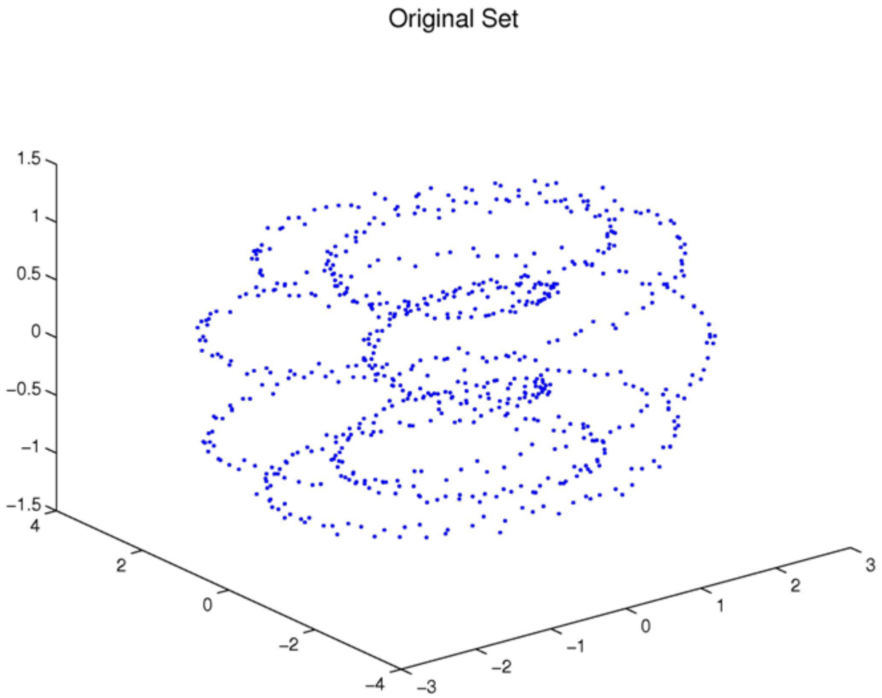
$\Rightarrow$ **dimensionality reduction of high dimensional data**

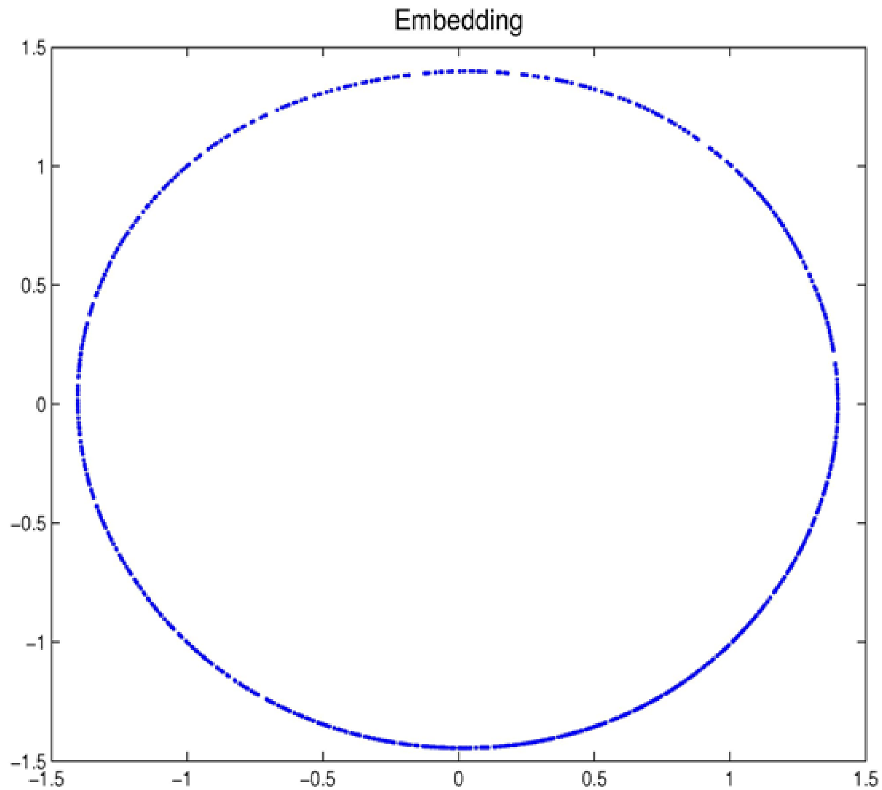

Original Set
1
0.5
0
−0.5
−1
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

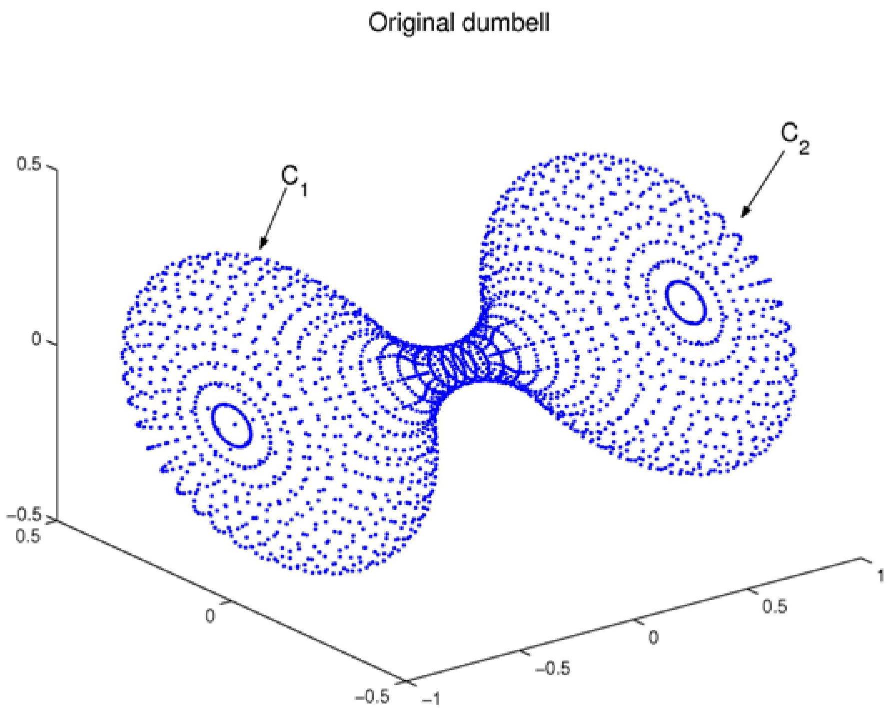

Embedding
1.5
1
0.5
0
−0.5
−1
−1.5
−1.5
−1
−0.5
0
0.5
1
1.5

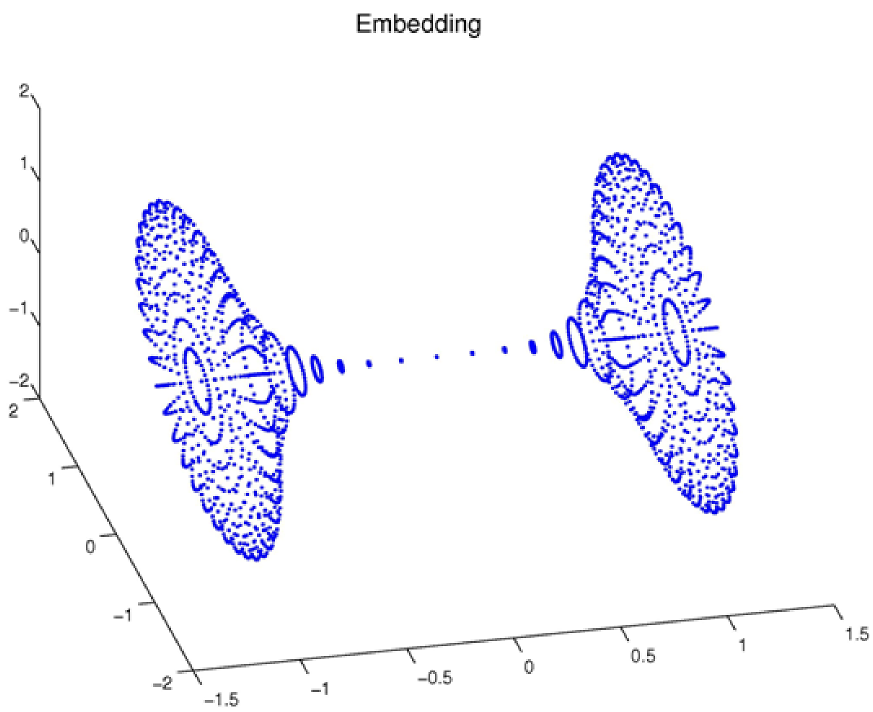

Trefoil
15
10
5
0
-5
-10
-15
-20
-10
0
10
20
30
30
20
10
0
-10
-20
-30


Embedding
1.5
1
0.5
0
−0.5
−1
−1.5
−1.5
−1
−0.5
0
0.5
1
1.5


Original Set
1.5
1
0.5
0
−0.5
−1
−1.5
4
2
0
−2
−4
−3
−2
−1
0
1
2
3


Embedding
1.5
1
0.5
0
−0.5
−1
−1.5
−1.5
−1
−0.5
0
0.5
1
1.5


Original dumbell
C1
C2
0.5
0
−0.5
0.5
0
−0.5
−1
−0.5
0
0.5
1


Embedding
2
1
0
−1
−2
2
1
0
−1
−2
−1.5
−1
−0.5
0
0.5
1
1.5

(d) Applications

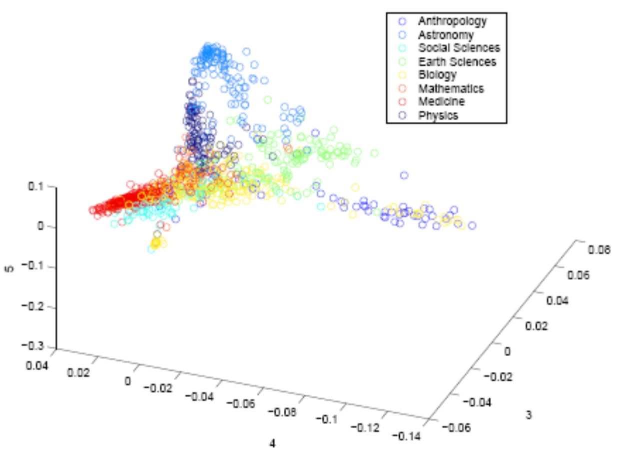
Spectral methods:

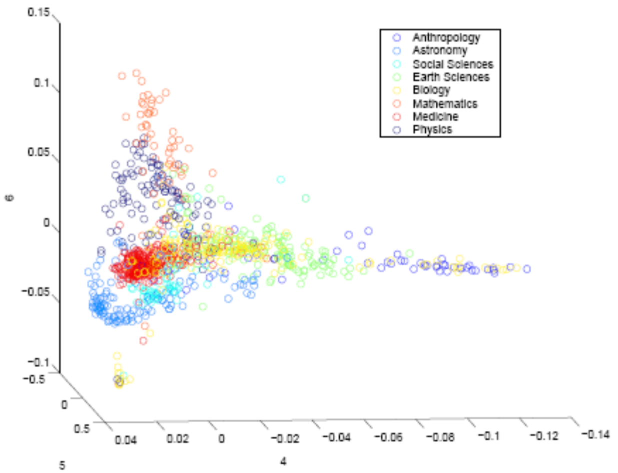
- start from the knowledge of the local geometry and infers a global representation
- allow a non-linear re-organizing and dimension reduction of data sets (graphs and more general manifolds)
- are well-suited for subsequent tasks s.a. visualization, clustering and partitioning of data

- Visualization (Coifman et. al.)

An example of a text document corpus

Consider about 1000 Science News articles, from 8 different categories. For each we compute about 10000 coordinates, the i-th coordinate of document d representing the frequency in document d of the i-th word in a fixed dictionary. The diffusion map gives the embedding below.





Embedding $\Xi_6^{(0)}(x) = (\xi_1(x), \ldots, \xi_6(x))$: on the left coordinates $3, 4, 5$, and on the right coordinates $4, 5, 6$.

- Segmentation via normalized cuts

J. SHI, J. MALIK, NORMALIZED CUTS AND IMAGE SEGMENTATION,
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO.8, AUGUST 2000
M. MEILA, J. SHI, A RANDOM WALKS VIEW OF SPECTRAL SEGMENTATION,
AISTATS 2001

Task:

partition a weighted graph $G = (V, E, W)$ into disjoint sets A, B, such that

$$\mathrm{Ncut}(A, B) = \mathrm{cut}(A, B) \left(\frac{1}{\mathrm{assoc}(A, V)} + \frac{1}{\mathrm{assoc}(B, V)} \right)$$

(where $\mathrm{cut}(A, B) = \sum_{u \in A, v \in B} W(u, v)$ and $\mathrm{assoc}(A, V) = \sum_{u \in A, v \in V} W(u, v)$)
is minimized.

'find a cut of relatively small weight between two subsets with strong internal connection'

Minimization of NCut is NP-hard.

Ncut algorithm:

Approximate method of solving the Ncut problem using a **spectral method**:

Calculate Eigenvalues and Eigenvectors of the Laplacian matrix, use Eigenvector of the second smallest Eigenvalue to bipartition the graph.

Random walk point of view:
Small Eigenvalues of the Laplacian are (up to normalization) the large Eigenvalues of the Diffusion matrix associated to a random walk
$\Rightarrow$ Partitioning of the graph into two parts such that the random walk, once in one of the parts, tends to remain in it

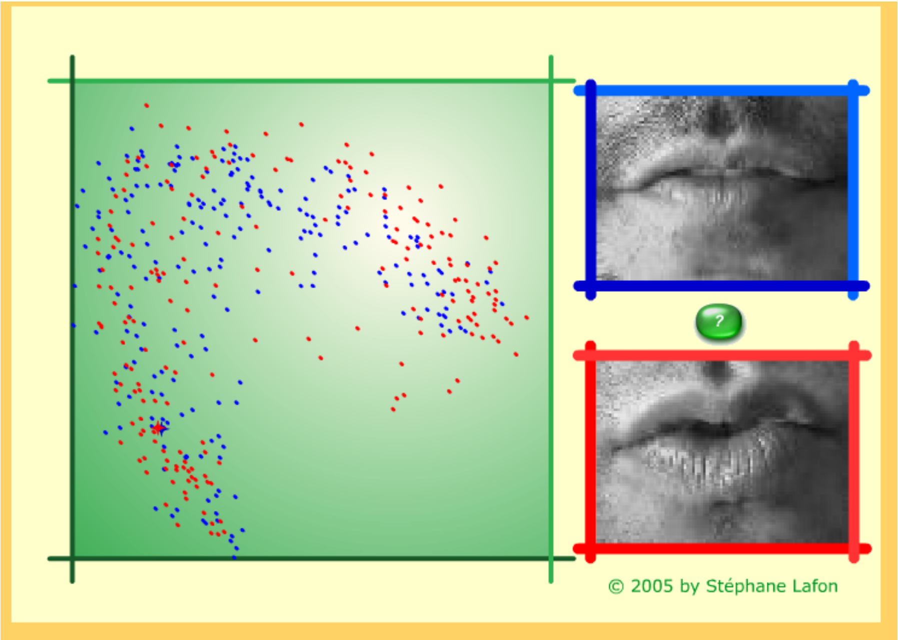
Fourier analysis point of view:
Functions f on a graph are projected on a subspace of low-frequency approximations (first term(s) of a Fourier series)
$\Rightarrow$ ideal low pass filtering

• Data matching

S. LAFON, Y. KELLER AND R.R. COIFMAN, DATA FUSION AND MULTI-CUE DATA MATCHING BY DIFFUSION MAPS, TO APPEAR IN IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

Given two data sets, both represented by a graph ⇒ instead of direct comparison, compare their embedding via diffusion maps

Example: lip reading



There are a lot of other applications (e.g. event detection in video: Zhong, Shi, Visontai; Porikli, Hada).

Spectral graph theory provides an abstract framework for all of these methods (global information from local similarities).

This abstract framework, especially the ‘Frequency point of view’ allows to go further: information on multiple scales through local similarities

$\Rightarrow$ Techniques from harmonic analysis allow

Wavelet (Time-Frequency) Analysis on a graph

2. Spectral Graph Theory - Multiscale Analysis

(a) Wavelets on $\mathbb{R}$

For engineers: Wavelet transform signal $x = (x_n)_{n\in\mathbb{Z}} \xrightarrow{\text{FWT}} (d_{j,k})_{j,k\in\mathbb{Z}}$,

$$\begin{array}{ccccccccccc} x = a_0 & \rightarrow & a_1 & \rightarrow & a_2 & \rightarrow & \dots & \rightarrow & a_{j-1} & \rightarrow & a_j \\ & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & \\ & & d_1 & & d_2 & & & & d_{j-1} & & d_j \end{array}$$

where each horizontal arrow represents the same filtering and subsampling step $a_{j+1} =\downarrow_2 (a_j * g)$, and similarly, $d_{j+1} =\downarrow_2 (a_j * h)$, $\quad g, h$ CMF.

$(a_{j,l})_l$ approximation coefficients at scale j,
$(d_{j,l})_l$ wavelet coefficients at scale j.

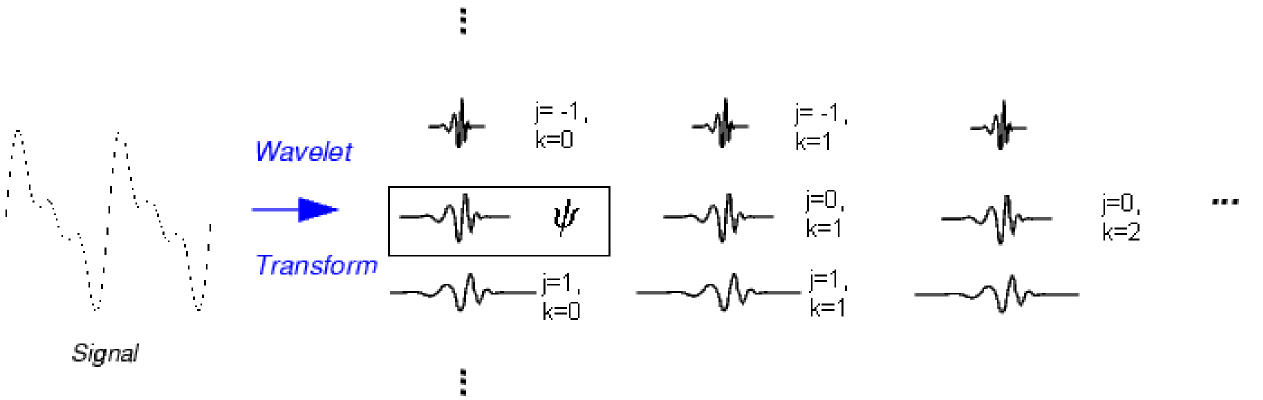
- Mathematical Formulation:

FWT corresponds to discretization of Wavelet Series

For $f \in L^2(\mathbb{R})$,

$$f = \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,l}\rangle \psi_{j,l},$$

$(\psi_{j,l})$ ONB in $L^2(\mathbb{R})$,
where $\psi_{j,l}$ are dilated and translated versions of a mother wavelet $\psi \in L^2(\mathbb{R})$.



Constituent wavelets of different scales and positions

Comparison to Fourier series:

- Fourier series
  - decomposition into sines and cosines having infinite support: frequency representation
  - global analysis
  - analysis of smooth functions (Sobolev class)
- Wavelet series
  - decomposition into localized functions at different scales: time-frequency representation
  - analysis on multiple scales
  - analysis of not-so-smooth functions (Besov class)

Wavelet algorithms are widely used for compression and denoising of images
Recipe: use largest coefficients for (non-linear!) approximation /+ Thresholding ('Wavelet Shrinkage'[Donoho])

• How to extend wavelet analysis on a graph?:

Dilations (stretching and squeezing) like on $\mathbb{R}$ cannot be used for scaling
$\Rightarrow$ use increasing powers of diffusion operators $(K^{2^j})_{j>0}$ as a scaling tool

• Formal framework: multiresolution analysis:

A multiresolution analysis (MRA) is a sequence of closed subspaces $(V_j)_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$, such that

$$\{0\} \subset \ldots \subset V_2 \subset V_1 \subset V_0 \subset \ldots L^2(\mathbb{R}),$$

there is $(\phi_{j,l})_{l\in\mathbb{Z}} \in V_j,$ *which is an orthonormal basis for* V_j.

$(\phi_{j,l})_l$ is called the family of scaling functions for (V_j).

In this setting (functions on $\mathbb{R}$), $\phi_{j,l} = (2^{-j/2}\phi(2^{-j}\cdot -l)))_{j,l\in\mathbb{Z}}$, where $\phi \in \mathcal{L}^2(\mathbb{R})$ is called **the** scaling function.

• But where are the wavelets?:

Define the spaces W_j by

$$V_{j-1} = V_j \oplus W_j.$$

An L^2 function ψ is called **wavelet**, if

$$(\psi_{j,l})_{l\in\mathbb{Z}} = (2^{-j/2}\psi(2^{-j}\cdot -l))_{l\in\mathbb{Z}}$$

is an orthonormal basis for W_j.
In addition, we have

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j$$

and the system $(\psi_{j,l})_{j,l\in\mathbb{Z}}$ constitutes an orthonormal basis for $L^2(\mathbb{R})$.

Projections on the spaces V_j: approximations of a function f at different resolutions,
partial wavelet series: difference between two approximation levels.

• Multiresolution Analysis on a Graph: Diffusion Wavelets [Coifman, Maggioni et al.]

Diffusion operator K as dilation operator acting on functions on $L^2(G)$
$\Rightarrow$ define multiresolution structure

$\{\lambda_i\}_{i\geq 0}$ (decreasingly ordered) spectrum of K with eigenvectors $\{\xi_i\}$,
$0 < \varepsilon < 1$, $t_j := 2^{j+1} - 1$, $j \geq 0$.

Divide the spectrum into 'low-pass' portions:

$$\sigma_j(K) := \{\lambda \in \sigma(K) : \lambda^{t_j} \geq \varepsilon\},$$

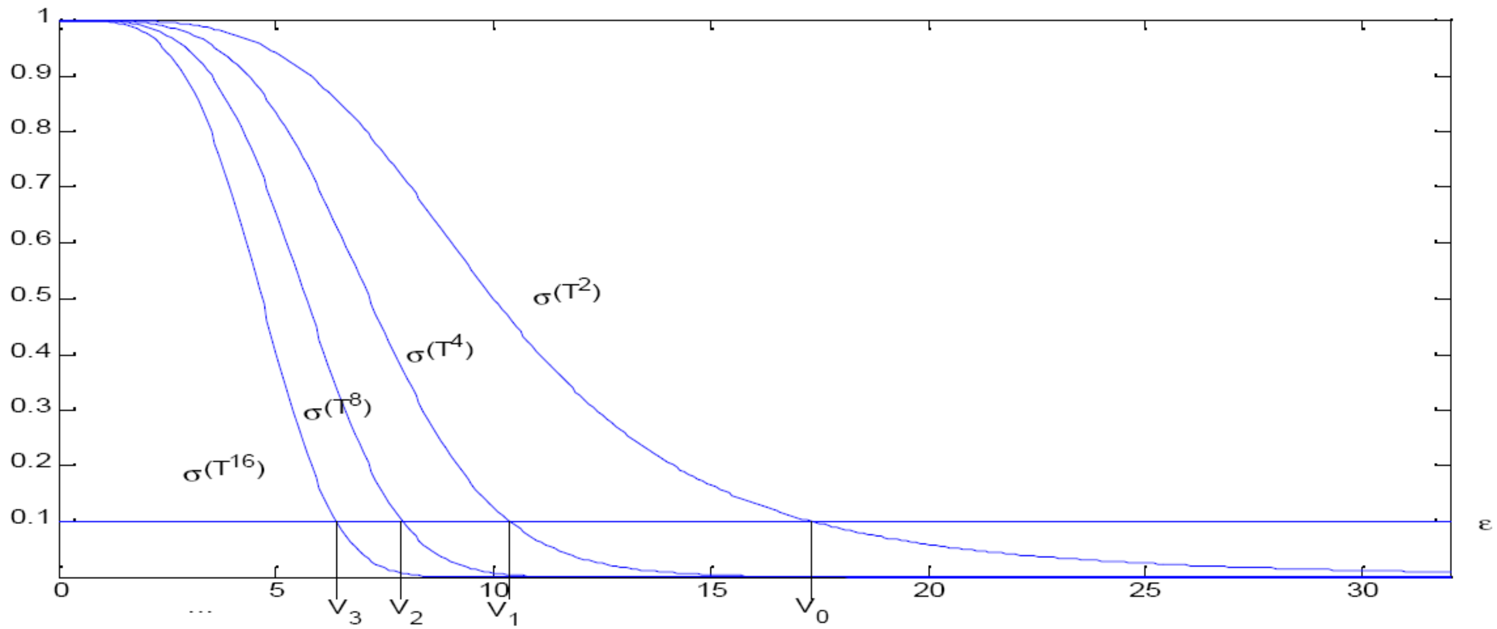
for $j \geq 0$ define the approximation spaces by

$$V_j := span\{\xi_\lambda : \lambda \in \sigma_j(K)\}$$

and wavelet spaces by $V_{j-1} = V_j \oplus W_j$.

$\Rightarrow$ Build localized orthonormal bases for V_j, W_j

Possible because for K with a fast decaying spectrum:
rank of K^{t^j} decreases $\Rightarrow$ compressibility



Algorithm:

$$\begin{tikzcd}
\Phi_0 \arrow[r, "M_0"] \arrow[d, "T_0"'] & \Phi_1 \arrow[r, "M_1"] \arrow[d, "T_1^2"'] & \cdots \arrow[r, "M_j"] & \Phi_{j+1} \arrow[r, "M_{j+1}"] \arrow[d, "T_j^{2^j}"'] & \cdots \\
\tilde{\Phi}_1 \arrow[ur, "G_0"] & \tilde{\Phi}_2 \arrow[ur, "G_1"] & \cdots \arrow[ur, "G_j"] & \tilde{\Phi}_{j+1} \arrow[ur, "G_{j+1}"] &
\end{tikzcd}$$

Fig. 1. Diagram for downsampling, orthogonalization and operator compression. (All triangles are commutative by construction)

$\Rightarrow$ Multiresolution analysis on graphs (in order $n(logn)^2$)

(Matlab code available on Mauro Maggioni's website)

With this construction: Calculating scaling and wavelet coefficients of functions on a graph (in order n).

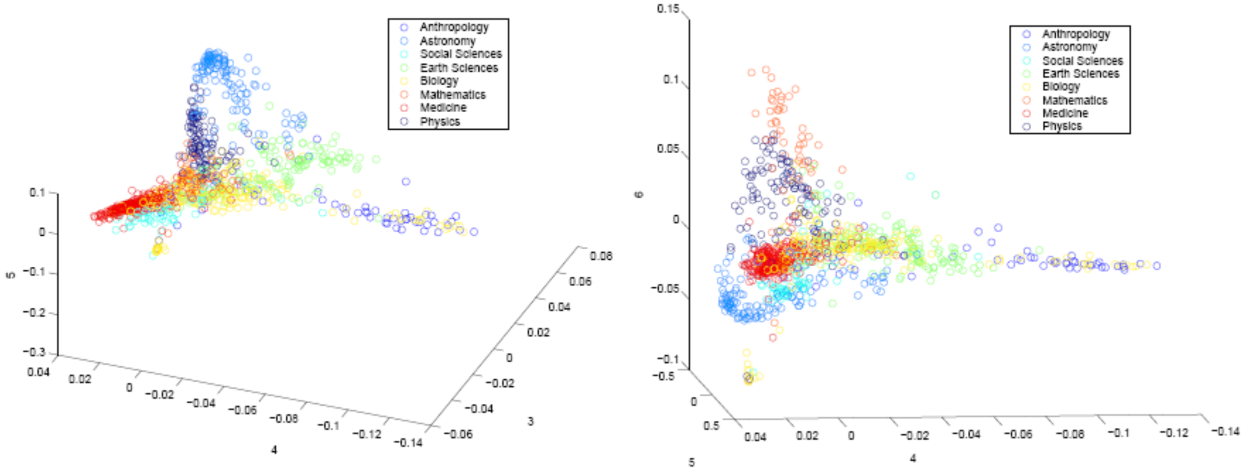
$\Rightarrow$ Multiscale transform in the spirit of classical wavelet analysis on non-linear structures (graphs, manifolds, general metric spaces)

(c) Examples

Back to the document problem

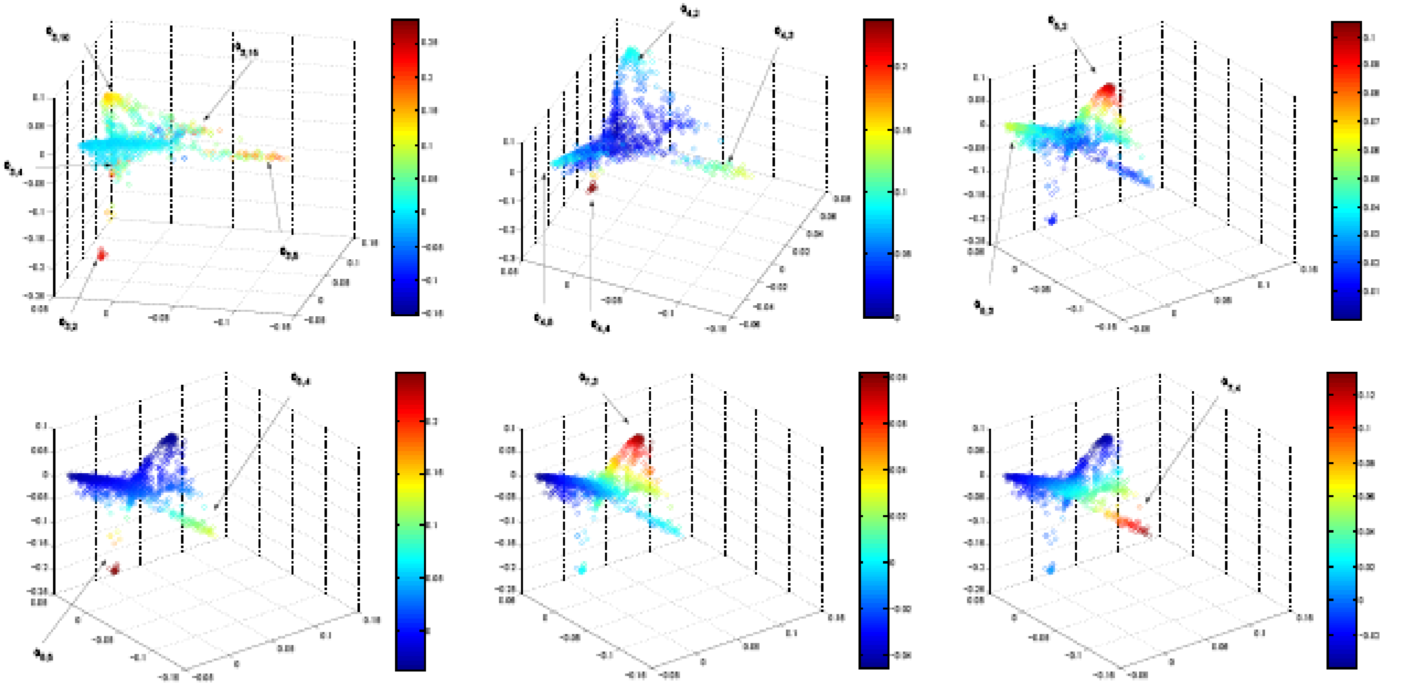
An example of a text document corpus

Consider about 1000 Science News articles, from 8 different categories. For each we compute about 10000 coordinates, the i-th coordinate of document d representing the frequency in document d of the i-th word in a fixed dictionary. The diffusion map gives the embedding below.



Embedding $\Xi_6^{(0)}(x) = (\xi_1(x), \ldots, \xi_6(x))$: on the left coordinates 3, 4, 5, and on the right coordinates 4, 5, 6.

Multiscale construction on a document corpus



Scaling functions at different scales represented on the set embedded in $\mathbb{R}^3$ via $(\xi_3(x), \xi_4(x), \xi_5(x))$. $\phi_{3,4}$ is about Mathematics, but in particular applications to networks, encryption and number theory; $\phi_{3,10}$ is about Astronomy, but in particular papers in X-ray cosmology, black holes, galaxies; $\phi_{3,15}$ is about Earth Sciences, but in particular earthquakes; $\phi_{3,5}$ is about Biology and Anthropology, but in particular about dinosaurs; $\phi_{3,2}$ is about Science and talent awards, inventions and science competitions.

Example of approximations on several scales of a planar domain with holes

[Szlam, Maggioni, Coifman, Bremer: Diffusion- driven Multiscale Analysis on Manifolds and Graphs: top-down and bottom-up constructions]

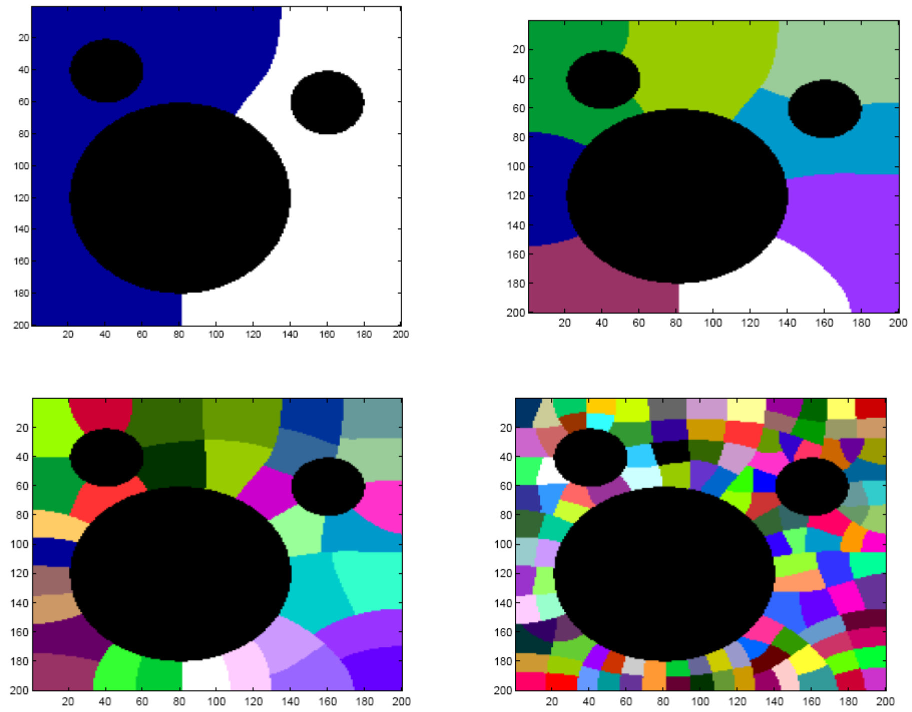


Figure 1. levels 2, 4, 6, and 8 in the decomposition of a planar domain with holes, uniformly sampled at 28,000 points.

3. Conclusion and POSSIBLE FURTHER APPLICATIONS

Spectral graph methods such as diffusion maps or diffusion wavelets:

- give abstract framework and tools for describing and processing non-linear (and maybe high dimensional) data structures such as graphs
- provide analogs to Fourier and wavelet based methods on these types of data
- connect harmonic analysis to a broad range of applied sciences, such as computer vision
- Possible applications:
  - Compression, structural complexity minimization, approximation; Denoising
  - (Multiscale) graph segmentation
  - (Multiscale) graph matching
  - (Multiscale) structure for learning tasks
  - ...

First concrete suggestions :

- Study the relationship between diffusion wavelet pyramids on a graph and graph pyramids (KRW et al.)
  $\Rightarrow$ first step: what does dual graph contraction to the eigenvalues of the Laplacian?
  How 'multiscale' is the algorithm?
  Derive hybrid constructions?

- Use the diffusion map embedding for matching images via graphs of feature points?

- Do this in a multiscale fashion using diffusion wavelets?

- Use the diffusion distance for the eccentricity transform (KRW et al.): Finding eccentric points on a surface by looking at their euclidean embedding via diffusion maps?
  (In this context, one should also study the embedding by Tenenbaum et.al (Science 2000) preserving geodesic distance.)