Multiscale harmonic analysis on graphs and data sets

Mauro Maggioni

Mathematics and Computer Science
Duke University

I.P.A.M., 09/23/08

In collaboration with R.R. Coifman, P.W. Jones, Y-M. Jung, R. Schul, A.D. Szlam & J.C. Bremer Jr.
Funding: NSF/DHS-FODAVA, DMS, IIS, CCF; ONR.

Plan

- Setting and Motivation
- Diffusion on Graphs
- Large time diffusion, eigenfunctions, spectral embeddings
- Multiscale analysis
  - Multiscale construction
  - Geometric and functional interpretation
  - Diffusion wavelets and algorithms
- Examples
- Conclusion

Structured data in high-dimensional spaces

A deluge of data: documents, web searching, customer databases, hyper-spectral imagery (satellite, biomedical, etc...), social networks, gene arrays, proteomics data, neurobiological signals, sensor networks, financial transactions, traffic statistics (automobilistic, computer networks)...
Common feature/assumption: data is given in a high dimensional space, however it has a much lower dimensional intrinsic geometry.

(i) physical constraints. For example the effective state-space of at least some proteins seems low-dimensional, at least when viewed at the time scale when important processes (e.g. folding) take place.

(ii) statistical constraints. For example many dependencies among word frequencies in a document corpus force the distribution of word frequency to low-dimensional, compared to the dimensionality of the whole space.

Structured data in high-dimensional spaces

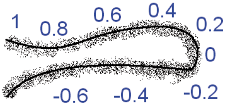
A deluge of data: documents, web searching, customer databases, hyper-spectral imagery (satellite, biomedical, etc...), social networks, gene arrays, proteomics data, neurobiological signals, sensor networks, financial transactions, traffic statistics (automobilistic, computer networks)...
Common feature/assumption: data is given in a high dimensional space, however it has a much lower dimensional intrinsic geometry.

(i) physical constraints. For example the effective state-space of at least some proteins seems low-dimensional, at least when viewed at the time scale when important processes (e.g. folding) take place.

(ii) statistical constraints. For example many dependencies among word frequencies in a document corpus force the distribution of word frequency to low-dimensional, compared to the dimensionality of the whole space.

Low-dimensional sets in high-dimensional spaces

In several instances the geometry of the data can help construct useful priors, for tasks such as classification, regression for prediction purposes.

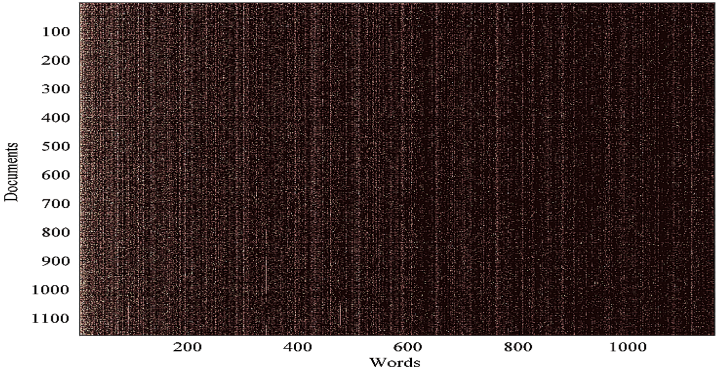


Some issues I am interested in:

- *geometric*: find intrinsic properties, such as local dimensionality, and local parameterizations.
- *approximation theory*: approximate functions on such data, respecting the geometry.

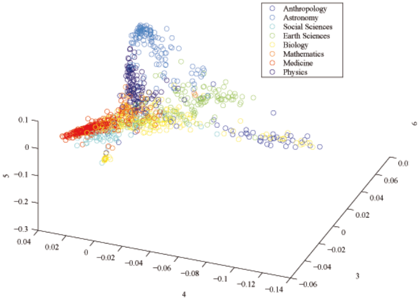
Text documents

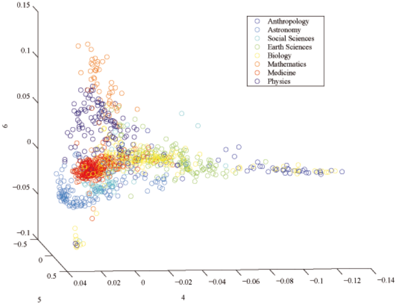
About 1100 Science News articles, from 8 different categories. We compute about 1000 coordinates, *i*-th coordinate of document *d* represents frequency in document *d* of the *i*-th word in a fixed dictionary.



Text documents

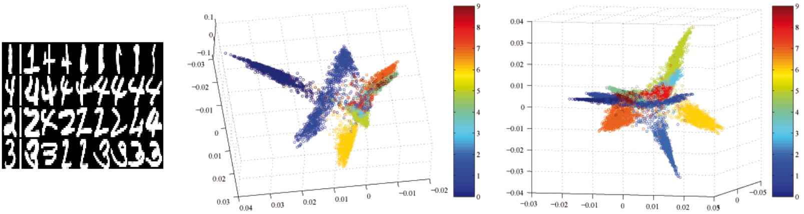
About 1100 Science News articles, from 8 different categories. We compute about 1000 coordinates, i-th coordinate of document d represents frequency in document d of the i-th word in a fixed dictionary.





Handwritten Digits

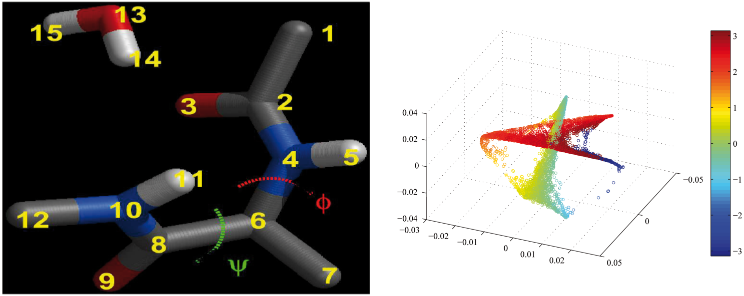
Data base of about $60,000$ 28×28 gray-scale pictures of handwritten digits, collected by USPS. Point cloud in R^{28^2}. Goal: automatic recognition.



Set of $10,000$ picture (28 by 28 pixels) of 10 handwritten digits. Color represents the label (digit) of each point.

[Joint with C. Clementi]

The dynamics of a small protein (12 atoms, H atoms removed) in a bath of water molecules is approximated by a Langevin system of stochastic equations $\dot{x} = -\nabla U(x) + \dot{w}$. The set of states of the protein is a noisy ($\dot{w}$) set of points in $\mathbb{R}^{36}$.



Left: representation of an alanine dipeptide molecule. Right: embedding of the set of configurations.

Goals

We start by analyzing the intrinsic geometry of the data, and then working on function approximation *on* the data.

- Find parametrizations for the data: manifold learning, dimensionality reduction. Ideally: number of parameters comparable with the intrinsic dimensionality of data + a parametrization should approximately preserve distances + be stable under perturbations/noise
- Construct useful dictionaries of functions on the data: approximation of functions on the manifold, predictions, learning.

Goals

We start by analyzing the intrinsic geometry of the data, and then working on function approximation *on* the data.

- Find parametrizations for the data: manifold learning, dimensionality reduction. Ideally: number of parameters comparable with the intrinsic dimensionality of data + a parametrization should approximately preserve distances + be stable under perturbations/noise
- Construct useful dictionaries of functions on the data: approximation of functions on the manifold, predictions, learning.

Random walks and heat kernels on the data

Assume the data $X = \{x_i\}_{i=1}^N \subset \mathbb{R}^D$. Assume we can assign local similarities via a kernel function $W(x_i, x_j) \geq 0$.
Simplest example: $W_\sigma(x_i, x_j) = e^{-||x_i - x_j||^2/\sigma}$.
Model the data as a *weighted graph* (G, E, W): vertices represent data points, edges connect x_i, x_j with weight $W_{ij} := W(x_i, x_j)$, when positive. Let $D_{ii} = \sum_j W_{ij}$ and

$$\underbrace{P = D^{-1}W}_{random\ walk}, \underbrace{T = D^{-\frac{1}{2}}WD^{-\frac{1}{2}}}_{symm.\ \text{``}random\ walk''}, \underbrace{H = e^{-tL}}_{Heat\ kernel}$$

Here $L = I - T$ is the normalized Laplacian.
Note 1: W depends on the type of data.
Note 2: W should be "local", i.e. close to 0 for points not sufficiently close.

Random walks and heat kernels on the data

Assume the data $X = \{x_i\}_{i=1}^N \subset \mathbb{R}^D$. Assume we can assign local similarities via a kernel function $W(x_i, x_j) \geq 0$.
Simplest example: $W_\sigma(x_i, x_j) = e^{-||x_i - x_j||^2/\sigma}$.
Model the data as a *weighted graph* (G, E, W): vertices represent data points, edges connect x_i, x_j with weight $W_{ij} := W(x_i, x_j)$, when positive. Let $D_{ii} = \sum_j W_{ij}$ and

$$\underbrace{P = D^{-1}W}_{random\ walk}, \quad \underbrace{T = D^{-\frac{1}{2}}WD^{-\frac{1}{2}}}_{symm.\ \text{``}random\ walk''}, \quad \underbrace{H = e^{-tL}}_{Heat\ kernel}$$

Here $L = I - T$ is the normalized Laplacian.
Note 1: W depends on the type of data.
Note 2: W should be "local", i.e. close to 0 for points not sufficiently close.

Random walks and heat kernels on the data

Assume the data $X = \{x_i\}_{i=1}^N \subset \mathbb{R}^D$. Assume we can assign local similarities via a kernel function $W(x_i, x_j) \geq 0$.
Simplest example: $W_\sigma(x_i, x_j) = e^{-||x_i - x_j||^2/\sigma}$.
Model the data as a *weighted graph* (G, E, W): vertices represent data points, edges connect x_i, x_j with weight $W_{ij} := W(x_i, x_j)$, when positive. Let $D_{ii} = \sum_j W_{ij}$ and

$$\underbrace{P = D^{-1}W}_{random\ walk}, \underbrace{T = D^{-\frac{1}{2}}WD^{-\frac{1}{2}}}_{symm.\ ``random\ walk''}, \underbrace{H = e^{-tL}}_{Heat\ kernel}$$

Here $L = I - T$ is the normalized Laplacian.
Note 1: W depends on the type of data.
Note 2: W should be "local", i.e. close to 0 for points not sufficiently close.

Basic properties

- $P^t(x, y)$ is the probability of jumping from x to y in t steps
- $P^t(x, \cdot)$ is a "probability bump" on the graph
- P and T are similar, therefore share the same eigenvalues $\{\lambda_i\}$ and the eigenfunctions are related by a simple transformation. Let $T\varphi_i = \lambda_i \varphi_i$, with $1 = \lambda_1 \geq \lambda_2 \geq \dots$.
- $\lambda_i \in [-1, 1]$
- "typically" P (or T) is large and sparse, but its high powers are full and low-rank

Functions on graphs

Any function $f : G \to \mathbb{R}$ is a vector in R^N. Euclidean norm and inner product:

$$||f||_2^2 = \sum_{x \in G} |f(x)|^2 d(x) \ , \ \langle f, g \rangle = \sum_{x \in G} f(x) g(x) d(x)$$

Other choices are possible

A Laplacian L allows to introduce a notion of smoothness

$$\langle Lf, f \rangle = \sum_x \sum_{y \sim x} W(x,y) \left(\frac{f(x)}{\sqrt{d_x}} - \frac{f(y)}{\sqrt{d_y}} \right)^2 \sim \int_{\text{edges}} |\nabla f|^2 dW$$

Moreover,

$$\lambda_i(L) = \min_{f \perp \langle \varphi_1, \dots, \varphi_{i-1} \rangle} \frac{\langle Lf, f \rangle}{||f||^2}$$

Functions on graphs

Any function $f : G \to \mathbb{R}$ is a vector in R^N. Euclidean norm and inner product:

$$||f||_2^2 = \sum_{x \in G} |f(x)|^2 d(x) \ , \ \langle f, g \rangle = \sum_{x \in G} f(x)g(x)d(x)$$

Other choices are possible
A Laplacian L allows to introduce a notion of smoothness

$$\langle Lf, f \rangle = \sum_x \sum_{y \sim x} W(x, y) \left(\frac{f(x)}{\sqrt{d_x}} - \frac{f(y)}{\sqrt{d_y}} \right)^2 \sim \int_{\text{edges}} |\nabla f|^2 dW$$

Moreover,

$$\lambda_i(L) = \min_{f \perp \langle \varphi_1, \ldots, \varphi_{i-1} \rangle} \frac{\langle Lf, f \rangle}{||f||^2}$$

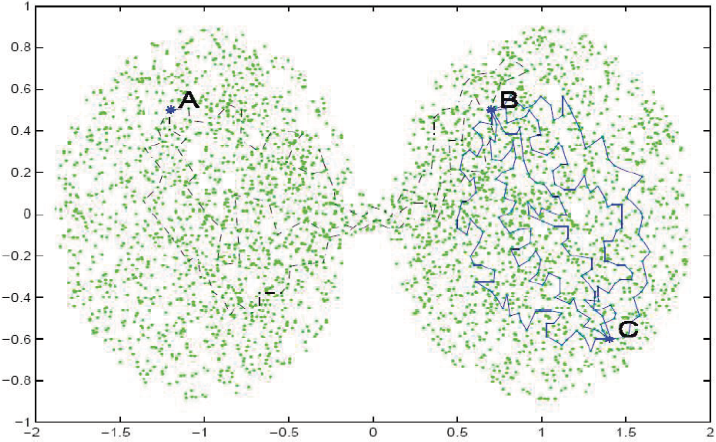
Dimensionality reduction and embeddings

Assume the data lies on a d-dimensional manifold $\mathcal{M}$ in $\mathbb{R}^n$ (think $n >> d$): how to find a map $\mathcal{M} \to \mathbb{R}^D$, with $D << n$ (hopefully $D \sim d$)?
Several techniques rely on a mapping

$$x \mapsto (\varphi_i(x))_{i=1,\dots,D} \,,$$

where φ_i are the eigenvectors of some matrix, e.g. dissimilarity matrix (CMDS), geodesic distance matrix (ISOMAP), or some local averaging operator (LLE, Laplacian eigenmap, Hessian eigenmap, etc...).
Pictures above: eigenfunctions of T: $T\varphi_i = \lambda_i \varphi_i$.

Diffusion distances



[Picture courtesy of S. Lafon]

Diffusion distances for large time

We would like to measure distances between points on a graph by random walks. Diffusion distance at time t:

$$\begin{aligned} d^{(2t)}(x,y) &= ||T^t\delta_x - T^t\delta_y|| = ||T^t(x,\cdot) - T^t(y,\cdot)|| \\ &= \sqrt{\sum_{z\in G} |T^t(x,z) - T^t(y,z)|^2} \\ &= \sqrt{\sum_i \lambda_i^t(\varphi_i(x) - \varphi_i(y))^2} \\ &\sim ||(\lambda_i^t\varphi_i(x))_{i=1}^m - (\lambda_i^t\varphi_i(y))_{i=1}^m||_{\mathbb{R}^m} \end{aligned}$$

Therefore $\Phi_m^{(2t)} : G \to \mathbb{R}^m$ with $\Phi_m^{(2t)}(x) = (\lambda_i^t\varphi(x))_{i=1}^m$ satisfies

$$||\Phi_m^{(2t)}(x) - \Phi_m^{(2t)}(y)||_{\mathbb{R}^m} \sim d^{(2t)}(x,y)$$

at least for t large and m large.

Analysis *on* the set

Equipped with good systems of coordinates on large pieces of the set, one can start doing analysis and approximation intrinsically on the set.

- *Fourier analysis on data*: use eigenfunctions for function approximation. Ok for globally uniformly smooth functions. Conjecture: most functions of interest are not in this class (Belkin, Niyogi, Coifman, Lafon).
- *Diffusion wavelets*: can construct multiscale analysis of wavelet-like functions on the set, adapted to the geometry of diffusion, at different time scales (joint with R.Coifman).
- The *diffusion semigroup* itself on the data can be used as a smoothing kernel. We recently obtained very promising results in image denoising and semisupervised learning (in a few slides, joint with A.D. Szlam and R. Coifman).

Applications

- Hierarchial organization of data and of Markov chains (e.g. documents, regions of state space of dynamical systems, etc...);
- Distributed agent control, Markov decision processes (e.g.: compression of state space and space of relevant value functions);
- Machine Learning (e.g. nonlinear feature selection, semisupervised learning through diffusion, multiscale graphical models);
- Approximation, learning and denoising of functions on graphs (e.g.: machine learning, regression, etc...)
- Sensor networks: compression of measurements collected from the network (e.g. wavelet compression on scattered sensors);
- Multiscale modeling of dynamical systems (e.g.: nonlinear and multiscale PODs);
- Compressing data and functions on the data;
- Data representation, visualization, interaction;
- ...

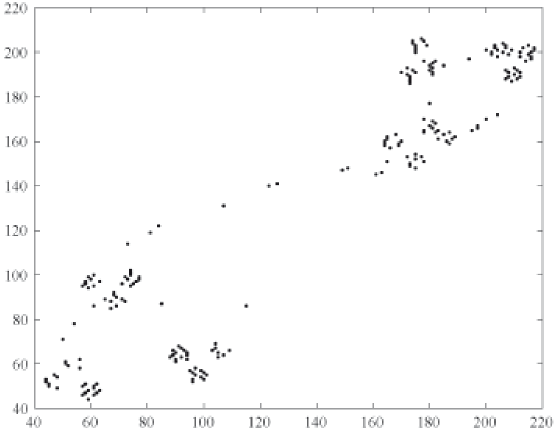
Summary for the “Fourier part”

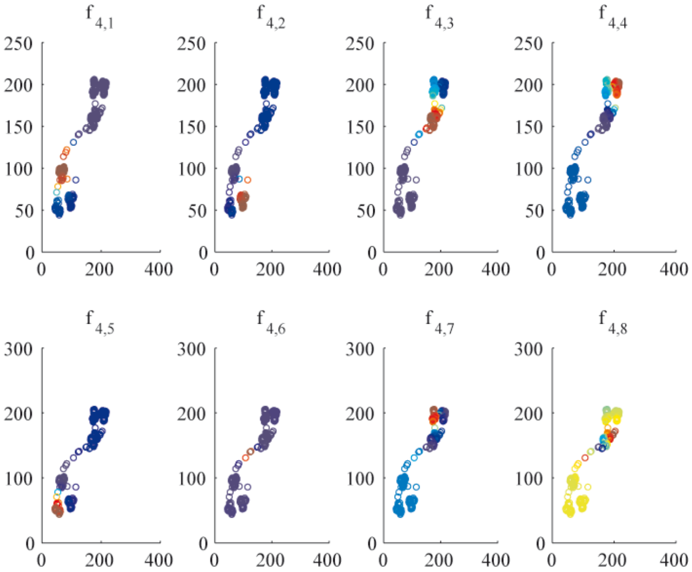
- it is useful to start with only local similarities between data points;
- it is possible to organize this local information by diffusion;
- parametrizations can be found by looking at the eigenvectors of a diffusion operator (Fourier modes);
- these eigenvectors yield a nonlinear embedding into low-dimensional Euclidean space;
- the eigenvectors can be used for global Fourier analysis on the set/manifold.

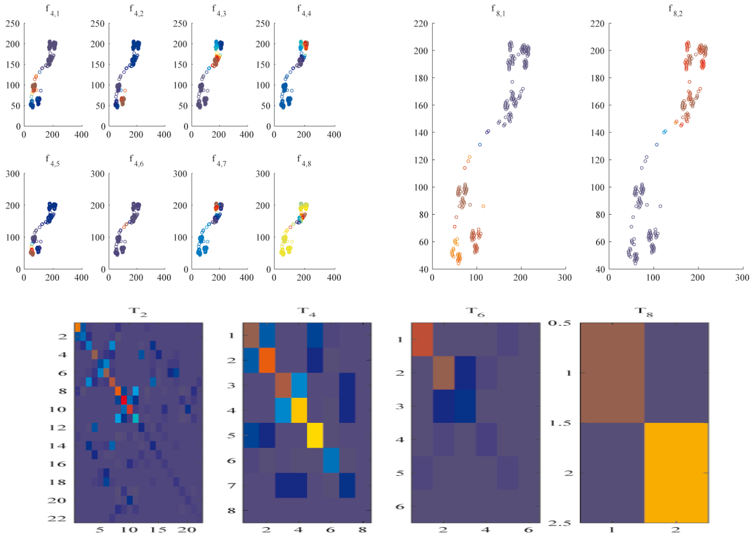
Next: going multiscale

Motivation: Either very local information or very global information: in many problems the intermediate scales are very interesting! Would like **multiscale** information!
Possibility 1: proceed *bottom-up*: repeatedly cluster together in a multi-scale fashion, in a way that is faithful to the operator: diffusion wavelets.
Possibility 2: proceed *top-bottom*: cut greedily according to global information, and repeat procedure on the pieces: recursive partitioning, local cosines...
Possibility 3: do *both*?

A multiscale "network"

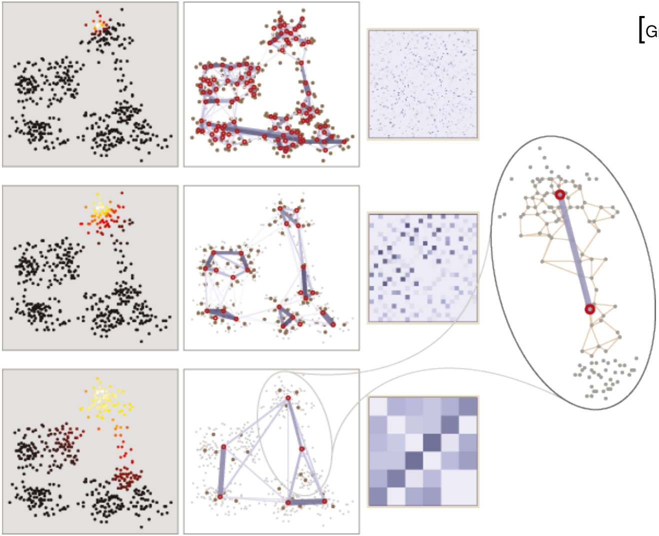



f 4,1
f 4,2
f 4,3
f 4,4
f 4,5
f 4,6
f 4,7
f 4,8
250
200
150
100
50
0
300
200
100
0
0
200
400



Multiscale elements and representation of powers of T

Multiscale Analysis, a sketch



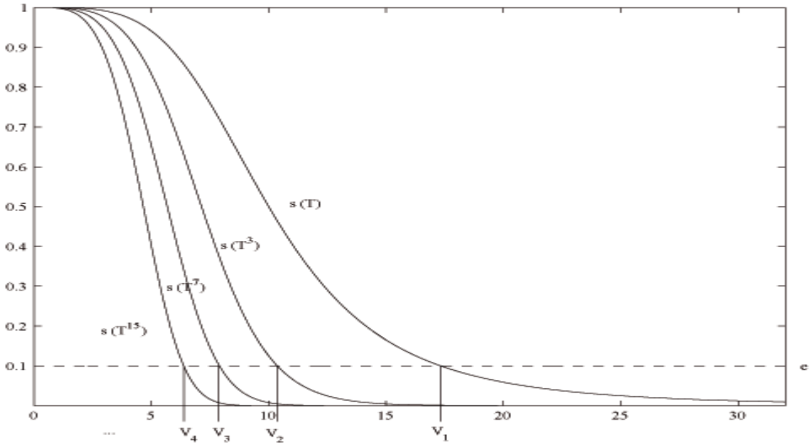
[Graphics by E. Monson]

Multiscale Analysis - what do we want?

We would like to be able to perform multiscale analysis *of* graphs, and of functions *on* graphs.
Of: produce coarser and coarser graphs, in some sense sketches of the original at different levels of resolution. This could allow a multiscale study of the geometry of graphs.
On: produce coarser and coarser functions on graphs, that allow, as wavelets do in low-dimensional Euclidean spaces, to analyse a function at different scales.
We tackle these two questions at once.

Multiscale Analysis, the spectral picture

Let $T = D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$ as above be the L^2-normalized symmetric "random walk".
The eigenvalues of T and its powers "typically" look like this:



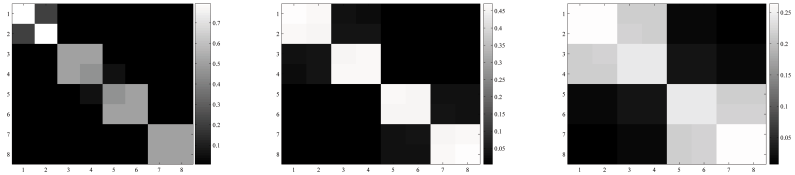
Multiscale Analysis, a trivial example, I

We now consider a simple example of a Markov chain on a graph with 8 states.

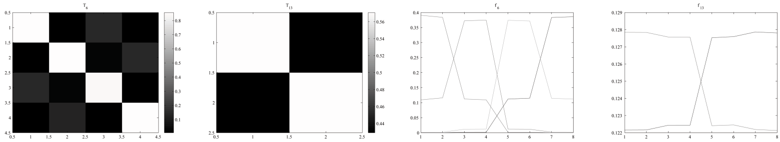
$$T = \begin{pmatrix} 0.80 & 0.20 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.20 & 0.79 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.49 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.50 & 0.499 & 0.001 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.001 & 0.499 & 0.50 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.49 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.01 & 0.49 & 0.50 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.50 \end{pmatrix}$$

From the matrix it is clear that the states are grouped into four pairs $\{\nu_1, \nu_2\}$, $\{\nu_3, \nu_4\}$, $\{\nu_5, \nu_6\}$, and $\{\nu_7, \nu_8\}$, with weak interactions between the the pairs.

Multiscale Analysis, a trivial example, II



Some powers of the Markov chain T, 8×8, of decreasing effective rank.



Compressed representations $T_6 := T^{2^6}$ (4×4), $T_{13} := T^{2^{13}}$ (2×2), and corresponding soft clusters.

Multiscale Analysis, a bit more precisely

We construct multiscale analyses associated with a diffusion-like process T on a space X, be it a manifold, a graph, or a point cloud. This gives:

(i) A coarsening of X at different "geometric" scales, in a chain $X \to X_1 \to X_2 \to \cdots \to X_j \ldots$;

(ii) A coarsening (or compression) of the process T at all time scales $t_j = 2^j$, $\{T_j = [T^{2^j}]_{\Phi_j}^{\Phi_j}\}_j$, each acting on the corresponding X_j;

(iii) A set of wavelet-like basis functions for analysis of functions (observables) on the manifold/graph/point cloud/set of states of the system.

All the above come *with guarantees*: the coarsened system X_j and coarsened process T_j have random walks "ϵ-close" to T^{2^j} on X. This comes at the cost of a very careful coarsening: up to $\mathcal{O}(|X|^2)$ operations ($< \mathcal{O}(|X|^3)$!), and only $\mathcal{O}(|X|)$ in certain special classes of problems.

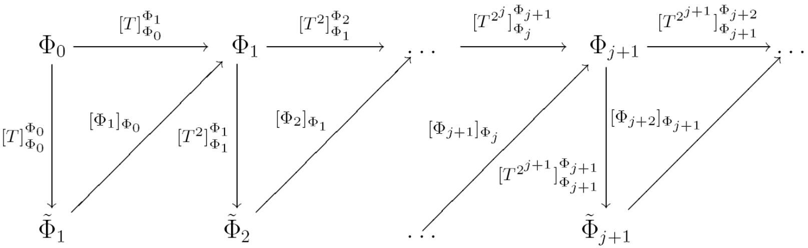


Figure: Diagram for downsampling, orthogonalization and operator compression. (All triangles are $\epsilon-$commutative by construction)

$\{\Phi_j\}_{j=0}^{J}, \{\Psi_j\}_{j=0}^{J-1}, \{[T^{2^j}]_{\Phi_j}^{\Phi_j}\}_{j=1}^{J} \leftarrow \texttt{DiffusionWaveletTree}\ ([T]_{\Phi_0}^{\Phi_0}, \Phi_0, J, \mathrm{SpQR}, \epsilon)$

// Input: $[T]_{\Phi_0}^{\Phi_0}$: a diffusion operator, written on the o.n. basis Φ_0
// Φ_0 : an orthonormal basis which ϵ-spans V_0
// J : number of levels to compute
// SpQR : a function compute a sparse QR decomposition, ϵ: precision

// Output: The orthonormal bases of scaling functions, Φ_j, wavelets, Ψ_j, representation of T^{2^j} on Φ_j.

for $j = 0$ **to** $J - 1$ **do**

$$[\Phi_{j+1}]_{\Phi_j}\ ,\ [T]_{\Phi_0}^{\Phi_1} \leftarrow \texttt{SpQR}([T^{2^j}]_{\Phi_j}^{\Phi_j}, \epsilon)$$

$$T_{j+1} := [T^{2^{j+1}}]_{\Phi_{j+1}}^{\Phi_{j+1}} \leftarrow [\Phi_{j+1}]_{\Phi_j}[T^{2^j}]_{\Phi_j}^{\Phi_j}[\Phi_{j+1}]_{\Phi_j}^*$$

$$[\Psi_j]_{\Phi_j} \leftarrow \mathrm{SpQR}(I_{\langle\Phi_j\rangle} - [\Phi_{j+1}]_{\Phi_j}[\Phi_{j+1}]_{\Phi_j}^*, \epsilon)$$

end

$Q, R \leftarrow \texttt{SpQR}\ (A, \epsilon)$
// Input: A: sparse $n \times n$ matrix ; ϵ: precision

// Output:
// Q, R matrices, possibly sparse, such that $A =_\epsilon QR$,
// Q is $n \times m$ and orthogonal,
// R is $m \times n$, and upper triangular up to a permutation,
// the columns of Q ϵ-span the space spanned by the columns of A.

Multiresolution Analysis

Let $V_j = \langle \Phi_j \rangle$, in fact Φ_j (scaling functions) is o.n. basis for V_j.
By construction $L^2(X) = V_0 \supseteq V_1 \supseteq V_2 \supseteq \dots$, and $V_j \to \langle \varphi_1 \rangle$.
Let W_j be the orthogonal complement of V_{j+1} into V_j. One can construct an o.n. basis Ψ_j (wavelets) for W_j.
$L^2(X) = W_0 \oplus \dots W_j \oplus V_j$, therefore we have

$$f = \sum_j \sum_{k \in \mathcal{K}_j} \underbrace{\langle f, \psi_{j,k} \rangle}_{\text{wavelet coeff.'s}} \psi_{j,k} \,.$$

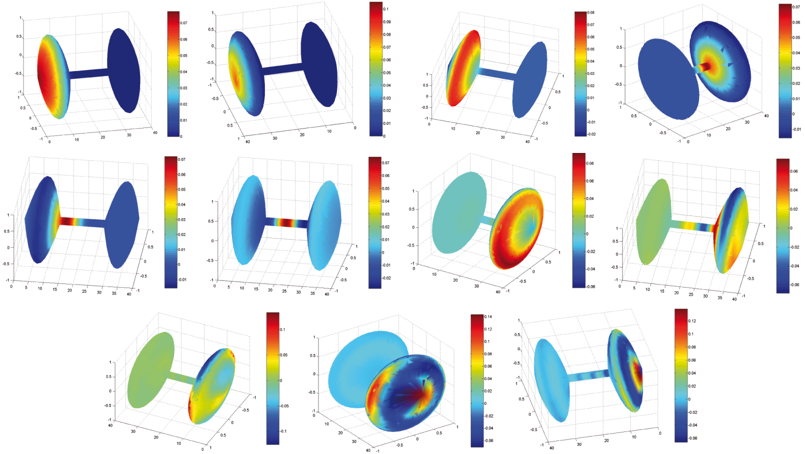
Signal processing tasks by adjusting wavelet coefficients.

Properties of Diffusion Wavelets

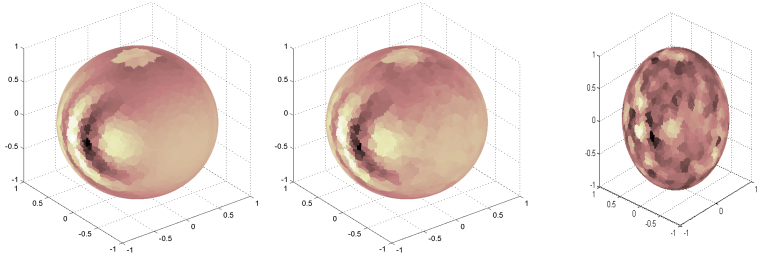
- Multiscale analysis and wavelet transform
- Compact support and estimates on support sizes (not as good as one really would like!);
- Vanishing moments (w.r.t. low-frequency eigenfunctions);
- Bounds on the sizes of the approximation spaces (depend on the spectrum of T, which in turn depends on geometry);
- Approximation and stability guarantees of the construction (tested in practice).

One can also construct diffusion wavelet packets, and therefore quickly-searchable libraries of waveforms.

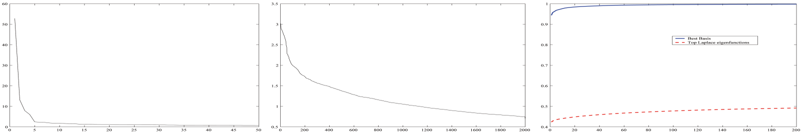
Diffusion Wavelets on Dumbell manifold



Signal Processing on Graphs

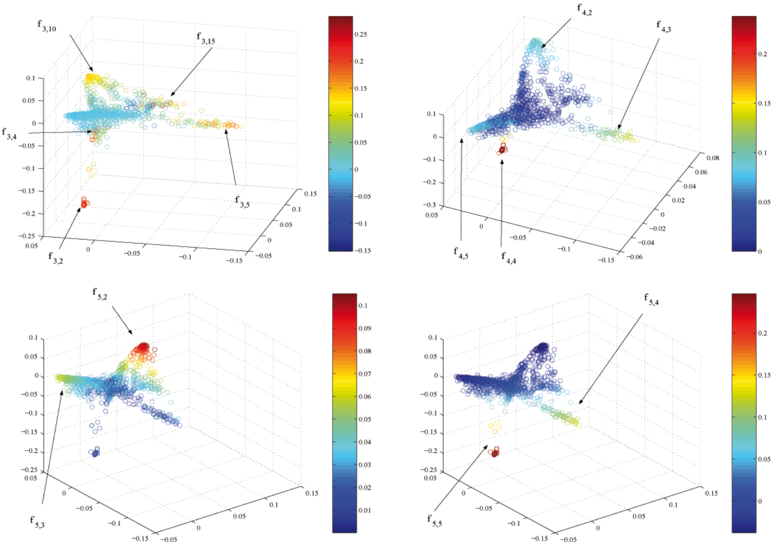


From left to right: function F; reconstruction of the function F with top 50 best basis packets; reconstruction with top 200 eigenfunctions of the Beltrami Laplacian operator.



Left to right: 50 top coefficients of F in its best diffusion wavelet basis, distribution coefficients$F$ in the delta basis, first 200 coefficients of F in the best basis and in the basis of eigenfunctions.

Example: Multiscale text document organization



Doc/Word multiscales

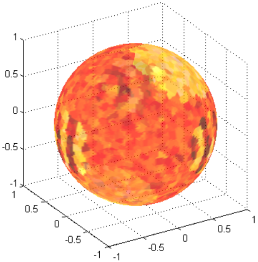
Scaling Fcn	Document Titles	Words
$\varphi_{2,3}$	Acid rain and agricultural pollution Nitrogen's Increasing Impact in agriculture	nitrogen,plant, ecologist,carbon, global
$\varphi_{3,3}$	Racing the Waves Seismologists catch quakes Tsunami! At Lake Tahoe? How a middling quake made a giant tsunami Waves of Death Seabed slide blamed for deadly tsunami Earthquakes: The deadly side of geometry	earthquake,wave, fault,quake, tsunami
$\varphi_{3,5}$	Hunting Prehistoric Hurricanes Extreme weather: Massive hurricanes Clearing the Air About Turbulence New map defines nation's twister risk Southern twisters Oklahoma Tornado Sets Wind Record	tornado,storm, wind,tornadoe, speed

Some example of scaling functions on the documents, with some of the documents in their support, and some of the words most frequent in the documents.

Local Discriminant Bases

One can in fact build a large dictionary of orthonormal bases (wavelet packets) by further splitting the wavelet subspaces into orthogonal subspaces.
Because of hierarchical organization, one can search such dictionary fast for “best bases” for tasks such as compression, denoising, classification.
LDB (Coifman, Saito) is the best basis for classification.

Local Discriminant Bases



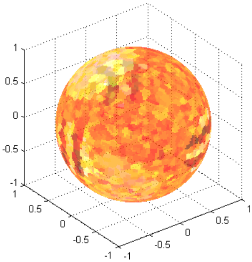


Figure: Left to right, a realization of a function from class 1 and 2 respectively. Note that the third smooth texture patch is on the back side of the sphere, and can be viewed in semitransparency. The other two smooth patches are decoys in random non-overlapping positions.

Some open questions and applications

Fourier part:

- Little known about global properties of eigenfunctions
- Behavior of eigenfunctions under perturbations of the graph
- Eigenfunctions on graphs different from sampled manifolds
- Relationships between eigenfunctions of different Laplacians

Multiscale part:

- Geometric multiscale properties of graphs
- Visualization of these multiscale decompositions
- Better constructions?

Applications

- Multiscale signal processing on graphs
- Multiscale learning and clustering on graphs
- We will see at least a couple of applications to the analysis of networks and network traffic in the next talks!

Acknowledgements

- R.R. Coifman, [Diffusion geometry; Diffusion wavelets; Uniformization via eigenfunctions; Multiscale Data Analysis], P.W. Jones (Yale Math), S.W. Zucker (Yale CS) [Diffusion geometry];
- P.W. Jones (Yale Math), R. Schul (UCLA) [Uniformization via eigenfunctions; nonhomogenous Brownian motion];
- S. Mahadevan (U.Mass CS) [Markov decision processes];
- A.D. Szlam (UCLA) [Diffusion wavelet packets, top-bottom multiscale analysis, linear and nonlinear image denoising, classification algorithms based on diffusion];
- G.L. Davis (Yale Pathology), R.R. Coifman, F.J. Warner (Yale Math), F.B. Geshwind , A. Coppi, R. DeVerse (Plain Sight Systems) [Hyperspectral Pathology];
- H. Mhaskar (Cal State, LA) [polynomial frames of diffusion wavelets];
- J.C. Bremer (Yale) [Diffusion wavelet packets, biorthogonal diffusion wavelets];
- M. Mahoney, P. Drineas (Yahoo Research) [Randomized algorithms for hyper-spectral imaging]
- J. Mattingly, S. Mukherjee and Q. Wu (Duke Math,Stat,ISDS) [stochastic systems and learning]; A. Lin, E. Monson (Duke Phys.) [Neuron-glia cell modeling]
- R. Brady, E. Monson (Duke CS) [Visualization]

Funding: NSF/DHS-FODAVA, DMS, IIS, CCF; ONR.

Thank you!

www.math.duke.edu/~mauro