Laplacian and wavelet bases for value function approximation and their connection to kernel methods

Mauro Maggioni

Department of Mathematics

Program in Applied Mathematics

Yale University

ICML Workshop, June 29th, 2006

Overview

1 - Diffusion & analysis on data sets;

2 - Multiscale analysis: bottom-up, top-bottom constructions;

3 - Applications to Markov Decision Processes;

4 - Applications to Semi-Supervised Learning;

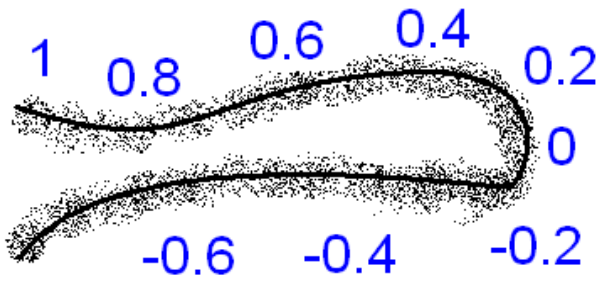
5 - Open problems, current and future research.

Big apology: large number of connections with the work of many researchers (in various fields of machine learning, harmonic analysis, PDEs, probability, image processing, numerical analysis), which I have no time to cite in this talk.

Parametrizations and functions on data sets

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 in many of these applications: data is given in "high-dimensional space", however it has "its own geometry" that is much lower dimensional. It is interesting to: discover and characterize intrinsic properties, such as local dimensionality, local parametrizations. Moreover, in many applications one needs to study functions on the data, and perform tasks such as approximation, smoothing, interpolation, extension to new data. One needs good basis functions on the data, with efficient algorithms for performing these operations.

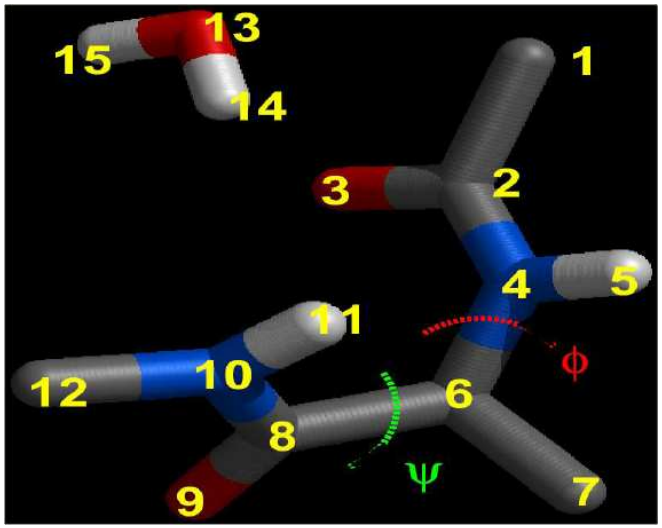


Regression "*on* the set"

An example from Molecular Dynamics

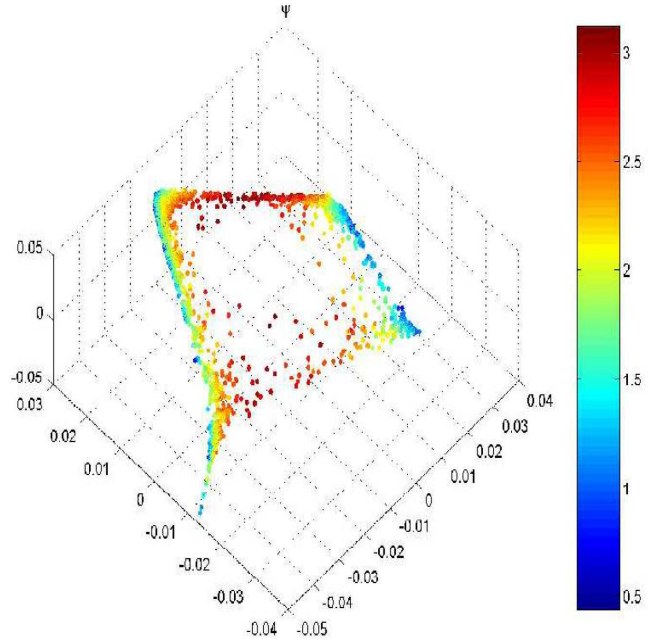
The dynamics of a small protein in a bath of water molecules is approximated by a Langevin system of stochastic equations $\dot{x} = -\nabla U(x) + \dot{w}$.

The set of states of the protein is a noisy set of points in $\mathbb{R}^{36}$, since we have 3 coordinates for each of the 12 atoms. This set is a priori very complicated. However we expect for physical reasons that the constraints on the molecule to force this set to be essentially lower-dimensional. We can explore the space of states by running long simulations, for different initial conditions.



The alanine molecule

In fact, as expected this set of points is much lower-dimensional.
We are able to discover this lower-dimensional
set, parametrize it, estimate
its local dimensionality (it is not constant!),
consider natural classes of (diffusion)
operators on this set, and build Fourier
and wavelet bases on it. For example it turns
out that some of the parameters discovered
by chemists on the basis of experiments
and chemical-physical considerations
can be discovered empirically as being
Fourier-like functions on the set of states!



Embedding of the set of states of the molecule.

The Heat Kernel and the Laplacian on Manifolds

Starting point: the **heat kernel** and **diffusion**(s).

In Euclidean space: Laplacian very natural because of its invariance under the natural symmetries of the space. Connections: Heat kernel - Brownian motion - potential theory - the heat equation.

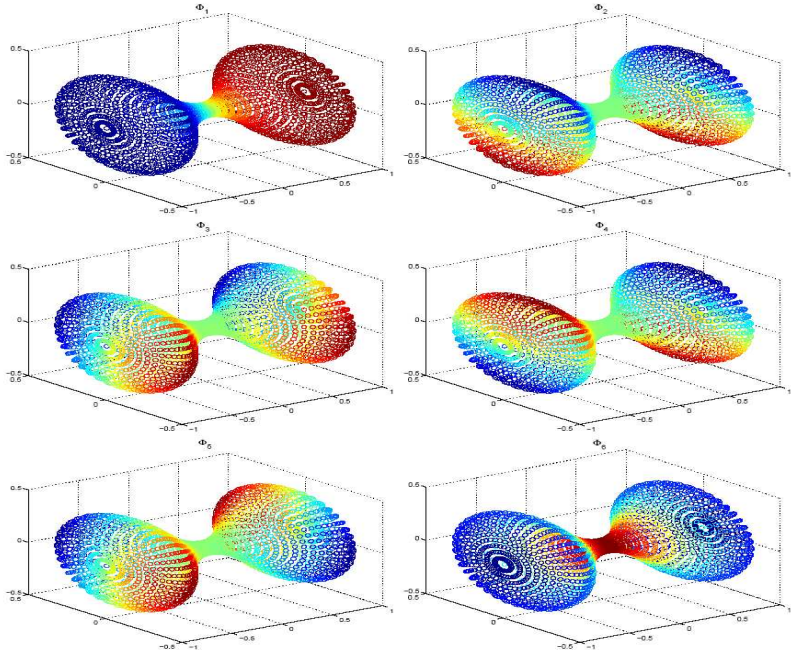
These objects are natural in much general settings, such as manifolds and graphs. Their properties relate to geometric properties of the space; they allow (for example through spectral theory) to define function spaces, operators, and bases that are natural generalization of their classical counterparts.s

On Riemannian manifolds we have the Laplace-Beltrami operator Δ.
Corresponding heat kernel $e^{-t\Delta}$ = Green's function for the heat equation on the manifold, associated with Brownian motion "restricted" to the manifold.
Spectral decomposition:

$$\Delta\phi_i = \lambda_i\phi_i \quad , \quad H_t(x,y) := e^{-t\Delta}(x,y) = \sum_i \underbrace{e^{-t\lambda_i}}_{\mu_i^t}\phi_i(x)\phi_i(y)\,.$$

The eigenfunctions ϕ_i of the Laplacian generalize Fourier modes: Fourier analysis on manifolds, global analysis.

Example of Eigenfunctions



Eigenfunctions on a dumbell-shaped manifold, and corresponding diffusion map; pictures courtesy of Stephane Lafon.

Rougher worlds: graph associated with data sets

A deluge of data: documents, web searching, customer databases, hyper-spectral imagery (satellite, biomedical, etc...), social networks, gene arrays, proteomics data, financial transactions, traffic statistics (automobilistic, computer networks)...

Assume we know how to assign local similarities: map data set to weighted graph. Global distances are not to be trusted!

Data often given as points in high-dimension, but constraints (natural, physical...) force it to be *intrinsically low-dimensional*.

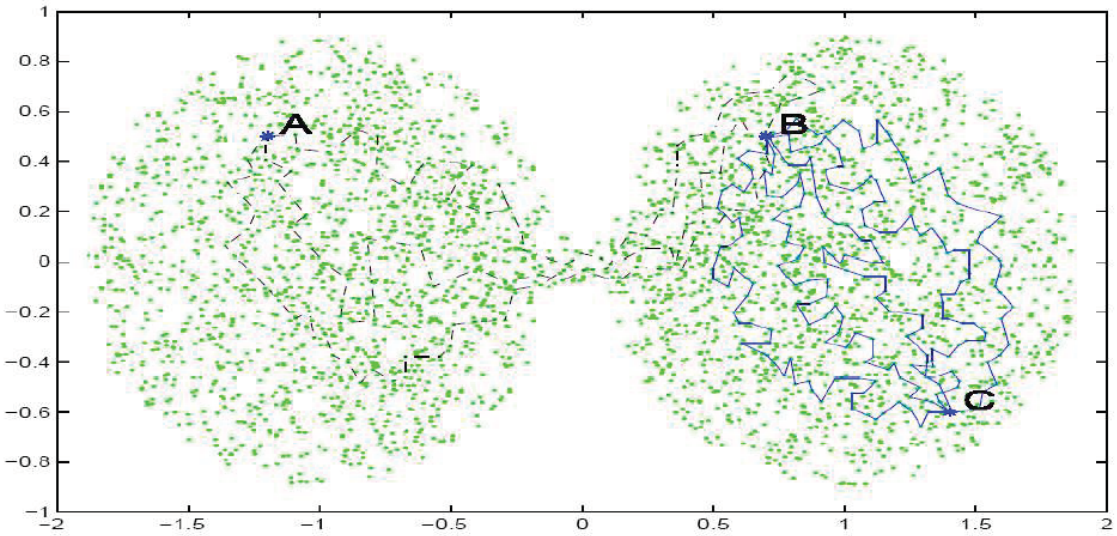
Model the data as a *weighted graph* (G, E, W): vertices represent data points (correspondence could be stochastic), edges connect similar data points, weights represent a similarity measure. Example: have an edge between web pages connected by a link; or between documents with very similar word frequencies. When points are in high-dimensional Euclidean space, weights may be a function of Euclidean distance, and/or the geometry of the points. How to define the similarity between very similar objects in each category is important but not always easy. That's the place where field-knowledge goes.

Laplacian on Graphs

Given a weighted graph (G, W, E), the combinatorial Laplacian is defined by $L = D - W$, where $(D)_{ii} = \sum_j W_{ij}$, and the normalized Laplacian is defined by

$$\mathcal{L} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}} .$$

These are self-adjoint positive-semi-definite operators, let λ_i and ϕ_i be the eigenvalues and eigenvectors. Fourier analysis on graphs. The heat kernel is of course defined by $H_t = e^{-t\mathcal{L}}$; the natural random walk is $D^{-1}W$.

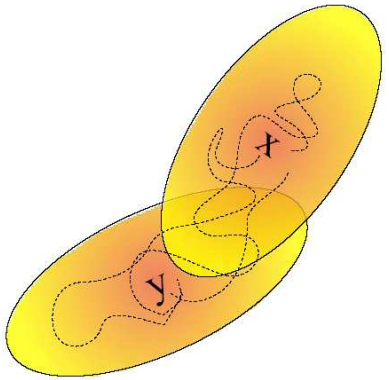


$d_{geod.}(A, B) \sim d_{geod.}(C, B)$, however $d^{(t)}(A, B) >> d^{(t)}(C, B)$.

Geometrization of Diffusion

Diffusion distance at time t is defined by

$$\begin{aligned} d^{(t)}(x,y) &= ||H_{t/2}(x,\cdot) - H_{t/2}(y,\cdot)||_{L^2(\mathcal{M})} \\ &= \sqrt{\langle \delta_x - \delta_y, H^t(\delta_x - \delta_y)\rangle} \\ &= \sqrt{\sum_i \mu_i^t (\phi_i(x) - \phi_i(y))^2} \end{aligned}$$



Surprisingly, the eigenfunctions of the Laplacian also allow to analyse the geometry of the manifold, and provide embeddings of the manifold ("diffusion maps"): for $m = 1, 2, \ldots$, $t > 0$ and $x \in \mathcal{M}$, define

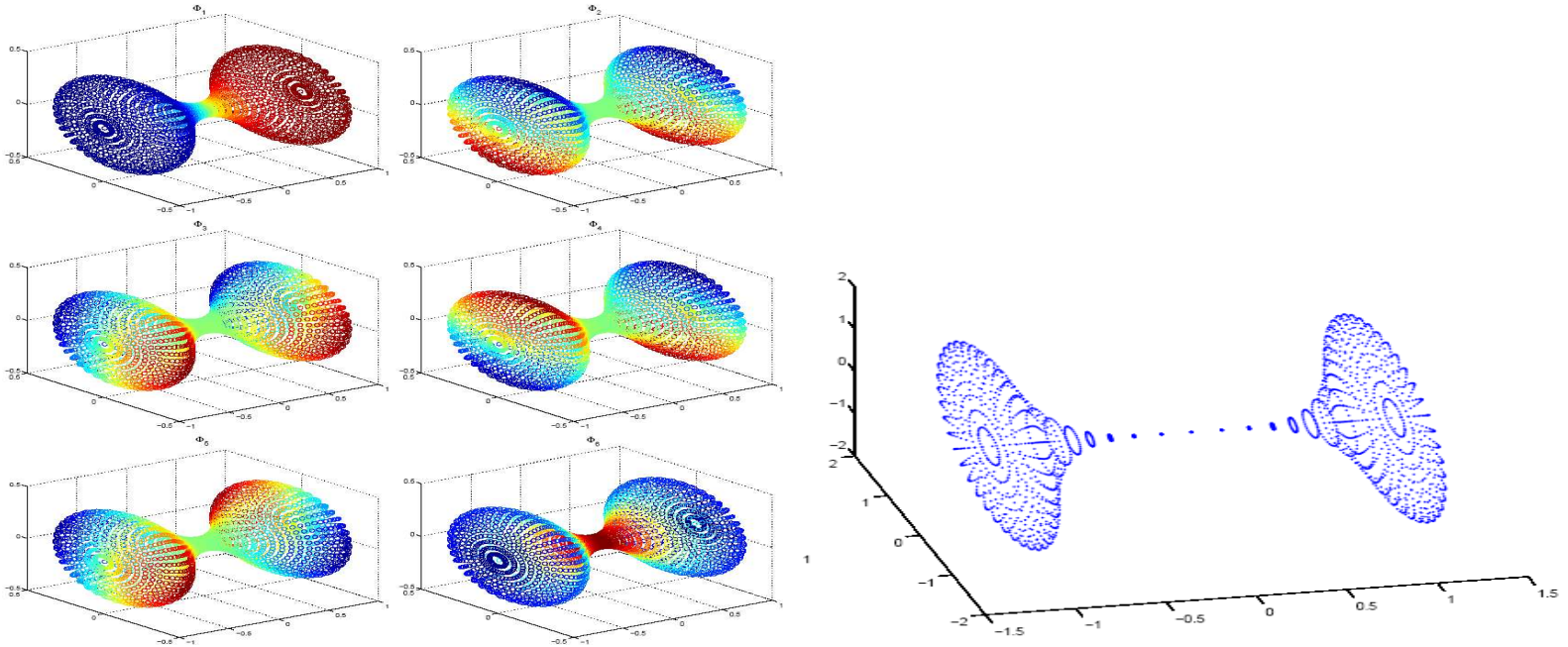
$$\Phi_m^{(t)}(x) = (\mu_i^{\frac{t}{2}} \phi_i(x))_{i=1,\ldots,m} \in \mathbb{R}^m .$$

This map is an approximate isometry (it is an isometry for $m = +\infty$) to Euclidean $\mathbb{R}^m$ from $\mathcal{M}$ with the diffusion metric (*not* the Riemannian metric!).

Diffusion maps: Example

Recall the definition of "diffusion map": for $m = 1, 2, \ldots,$ $t > 0$ and $x \in \mathcal{M}$, define

$$\Phi_m^{(t)}(x) = (\mu_i^{\frac{t}{2}} \phi_i(x))_{i=1,\ldots,m} \in \mathbb{R}^m .$$

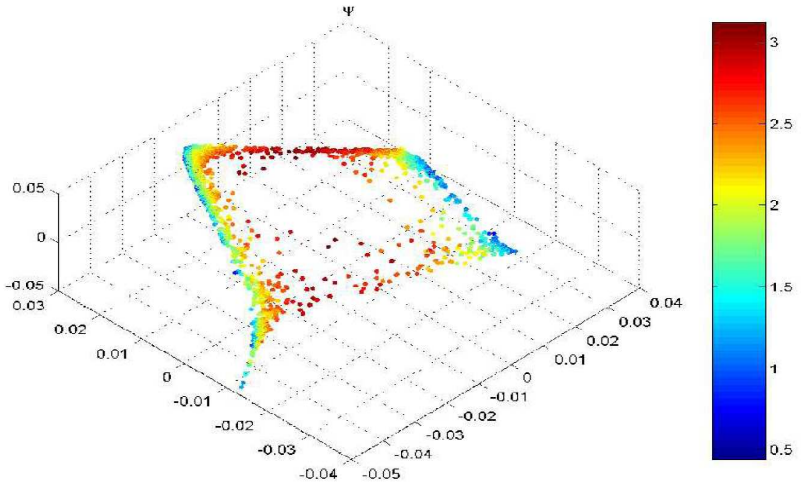


Eigenfunctions on a dumbell-shaped manifold, and corresponding diffusion map; pictures courtesy of Stephane Lafon.

Example from Molecular Dynamics revisited

The dynamics of a small protein in a bath of water molecules is approximated by a Langevin system of stochastic equations $\dot{x} = -\nabla U(x) + \dot{w}$. Many millions of points in $\mathbb{R}^{36}$ can be generated by simulating of the stochastic ODE, U is needed only "on the fly" and only at the current positions (not everywhere in $\mathbb{R}^{36}$).

Then a graph
Laplacian on this set of points can
be constructed, that approximated
the Fokker-Planck operator, and the
eigenfunctions of this approximation
yield a low-dimensional
description and parametrization
of the set, as well as a subspace
in which the long-term behavior
of the system can faithfully projected.



Embedding of the set of states of the molecule.

Stability of eigenfunctions, I

The data is often noisy, and the similarity matrices are usually noisy as well + useful to compare graphs of different sizes (e.g. graphs representing same physical system sampled at different rates).

Would like to have results that guarantee the stability of the eigenfunctions under rather general classes of transformations.

Model of perturbation: rough isometries. Let X, Y be two metric spaces. A map $\Phi : X \to Y$ is called a $(a'_\Phi, b'_\Phi, a_\Phi, b_\Phi, \tau_\Phi)$**-rough isometry**, for $a'_\Phi, b'_\Phi, \tau'_\Phi, a_\Phi, b_\Phi, \tau_\Phi \geq 0$, if for every $x_1, x_2 \in X$

$$a'_\Phi d_X(x_1, x_2) - b'_\Phi \leq d_Y(\Phi(x_1), \Phi(x_2)) \leq a_\Phi d_X(x_1, x_2) + b_\Phi$$

and for every $y \in Y$, $d_Y(y, \Phi(X)) \leq \tau_\Phi$.

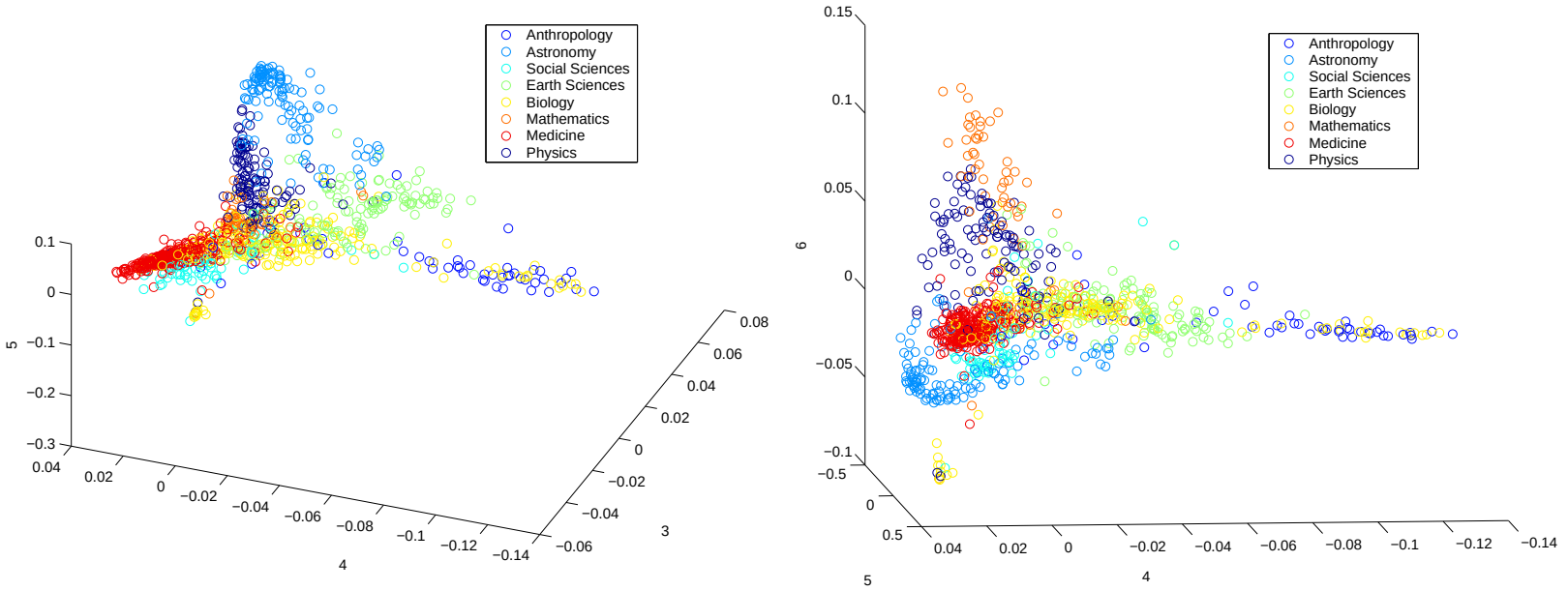
Several Applications

Many successful applications of spectral kernel methods. For Laplacian eigenfunctions, the following works in particular:

- Regression and classification in the supervised and semi-supervised learning context [M. Belkin, P. Nyogi; RR Coifman, MM, A.D. Slzam]
- fMRI data [F. Meyer, X. Shen]
- Art data [W Goetzmann, PW Jones, MM, J Walden]
- Hyperspectral Imaging in Pathology [MM, GL Davis, F Warner, F. Geshwind, A Coppi, R. DeVerse, RR Coifman]
- Molecular dynamics simulations [RR. Coifman, G.Hummer, I. Kevrekidis, S. Lafon, MM, B. Nadler]
- Text documents classification [RR. Coifman, S. Lafon, A. Lee, B. Nadler; RR Coifman, MM]

Application to text document classification

1000 Science News articles, from 8 different categories. We compute about 10000 coordinates, i-th coordinate of document d represents frequency in document d of the i-th word in a fixed dictionary. The diffusion map gives the embedding below. Clustering in the range of diffusion map results in good unsupervised performance for document classification.



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

Summary for the "Fourier part"

- it is useful to consider 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.

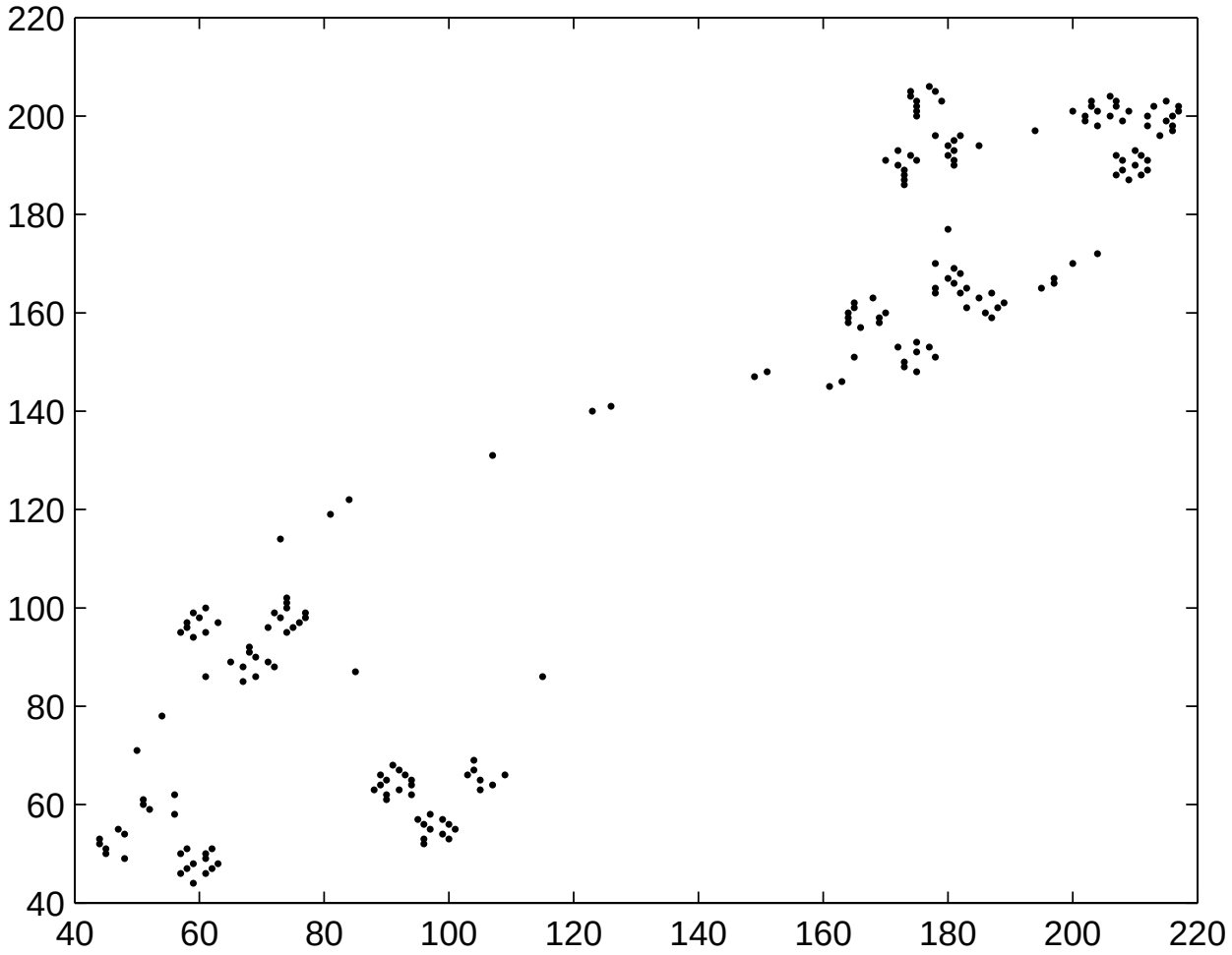
Problem: Either very local information or very global information: in many problems the intermediate scales are very interesting! Would like **multiscale** information!

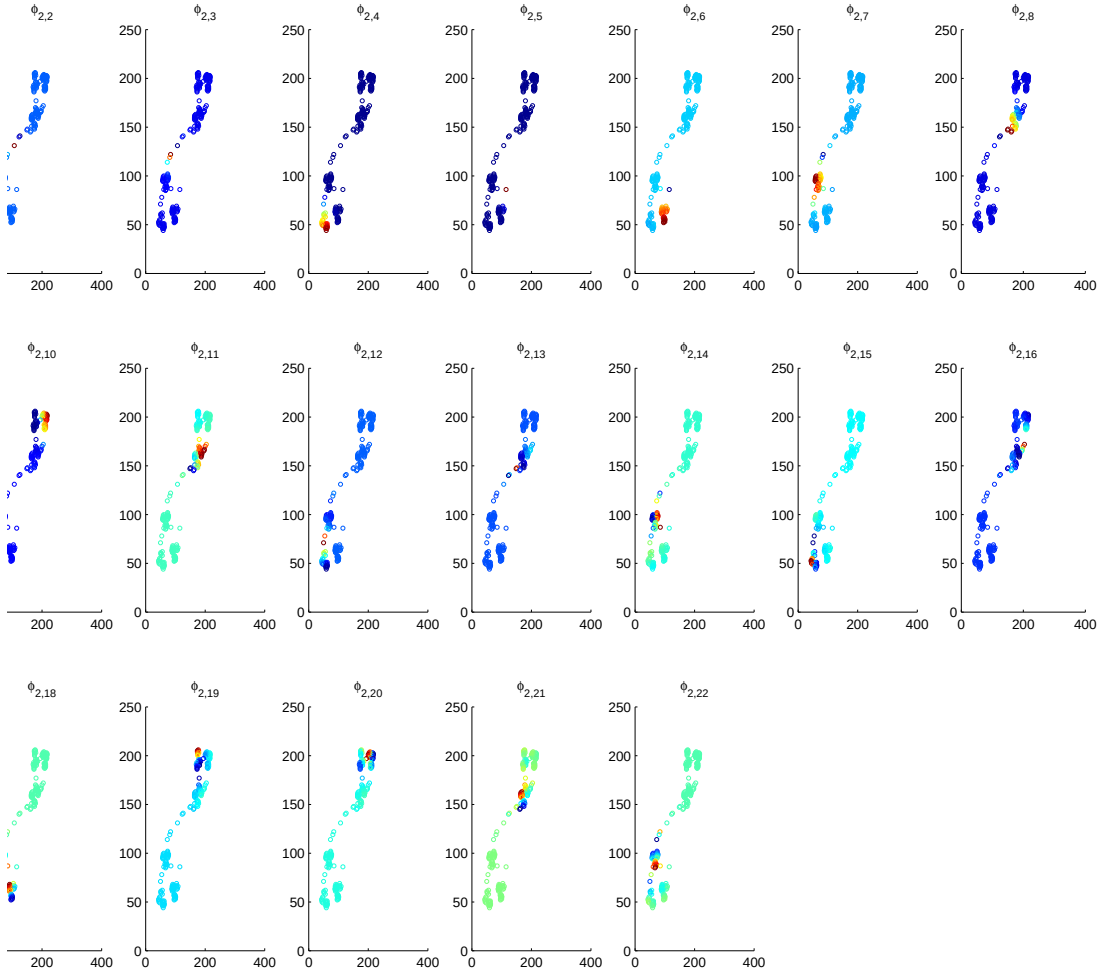
Possibility 1: proceed *bottom-up*: repeatedly cluster together in a multi-scale fashion, in a way that is faithful to the operator: diffusion wavelets.

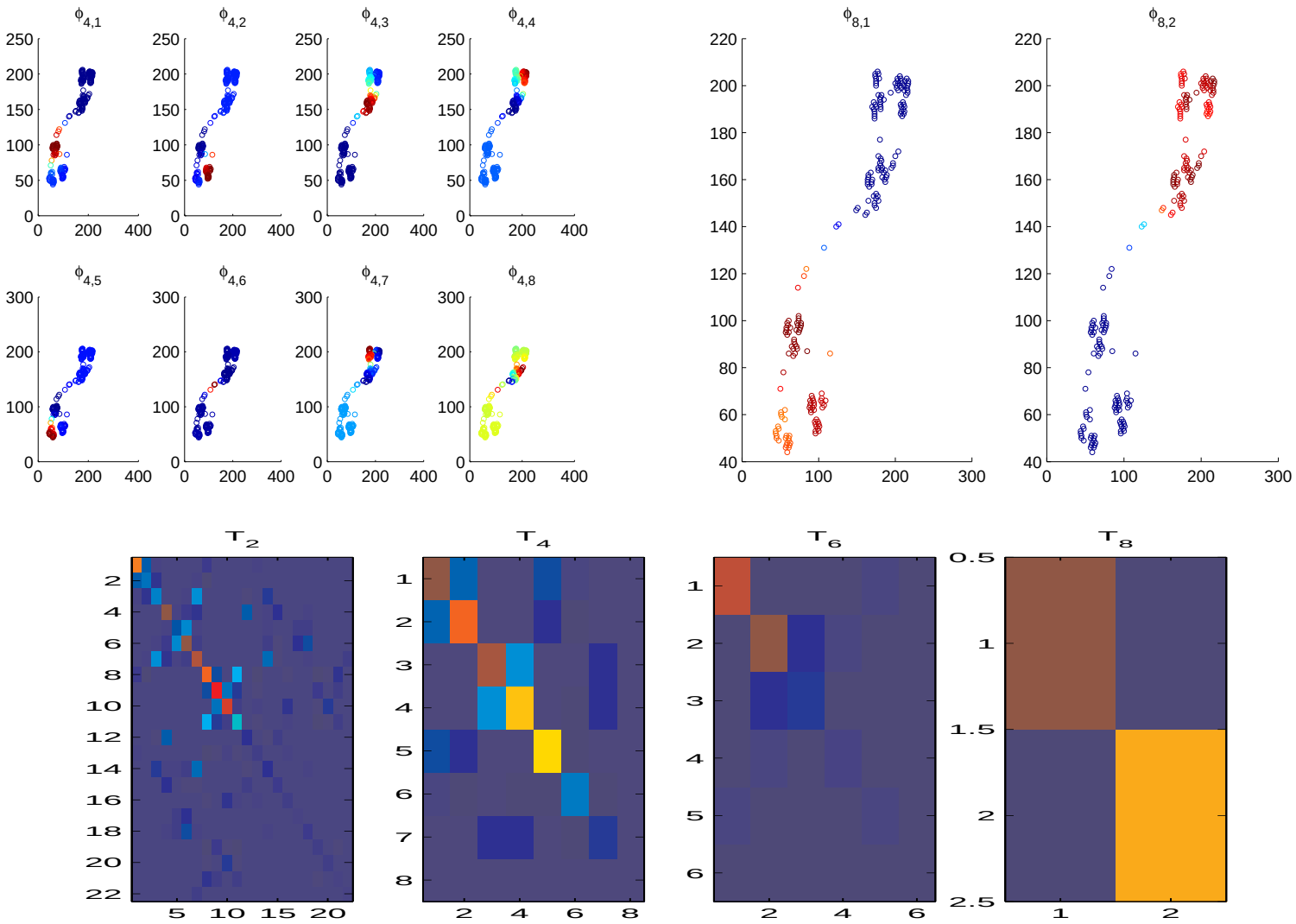
Possibility 2: proceed *top-bottom*: cut greedily according to global information, and repeat procedure on the pieces: recursive partitioning, local cosines...

Possibility 3: do *both*!

A multiscale "network"




$\phi_{2,2}$
$\phi_{2,3}$
$\phi_{2,4}$
$\phi_{2,5}$
$\phi_{2,6}$
$\phi_{2,7}$
$\phi_{2,8}$
$\phi_{2,10}$
$\phi_{2,11}$
$\phi_{2,12}$
$\phi_{2,13}$
$\phi_{2,14}$
$\phi_{2,15}$
$\phi_{2,16}$
$\phi_{2,18}$
$\phi_{2,19}$
$\phi_{2,20}$
$\phi_{2,21}$
$\phi_{2,22}$



Multiscale elements and representation of powers of T

Multiscale Analysis, I

As an analysis, we construct multiscale bases on manifolds, graphs, point clouds.

Classical constructions of wavelets are based on geometric transformations (such as dilations, translations) of the space, transformed into actions (e.g. via representations) on functions. There are plenty of such transformations on $\mathbb{R}^n$, certain classes of Lie groups and homogeneous spaces (with automorphisms that resemble "anisotropic dilations"), and manifolds with large groups of transformations.

Here the space is in general highly *non-symmetric*, not invariant under "natural" geometric transformation, and moreover it is *"noisy"*.

Idea: use *diffusion and the heat kernel as dilations*, acting on functions on the space, to generate multiple scales.

This is connected with the work on diffusion or Markov semigroups, and Littlewood-Paley theory of such semigroups (a la Stein).

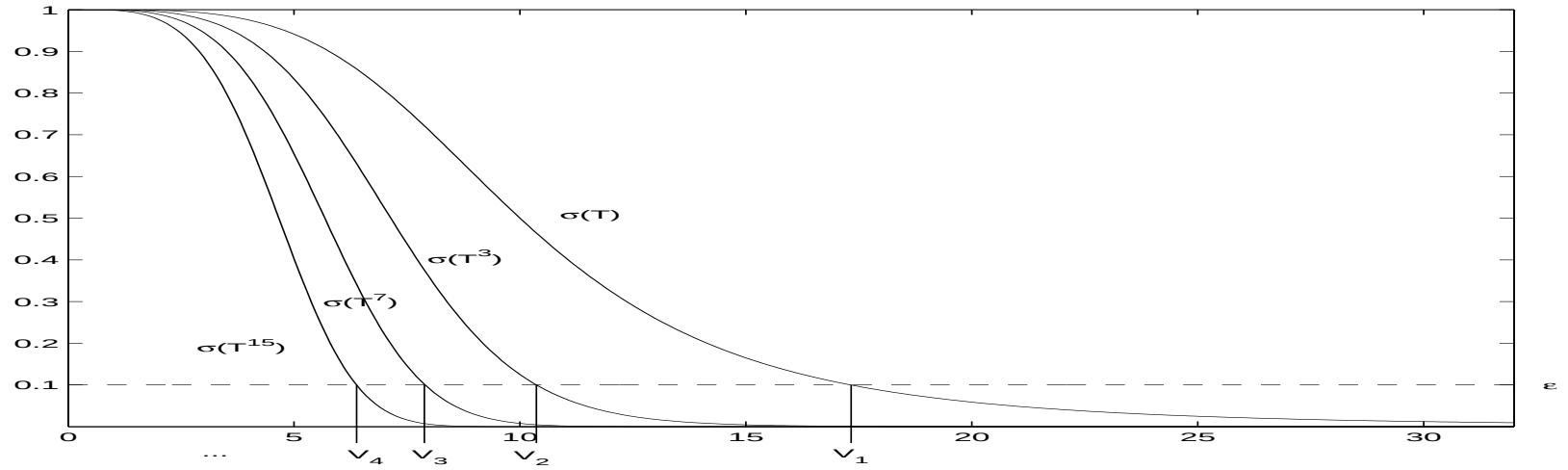
We would like to have *constructive* methods for efficiently computing the multiscale decompositions and the wavelet bases.

Multiscale Analysis, II

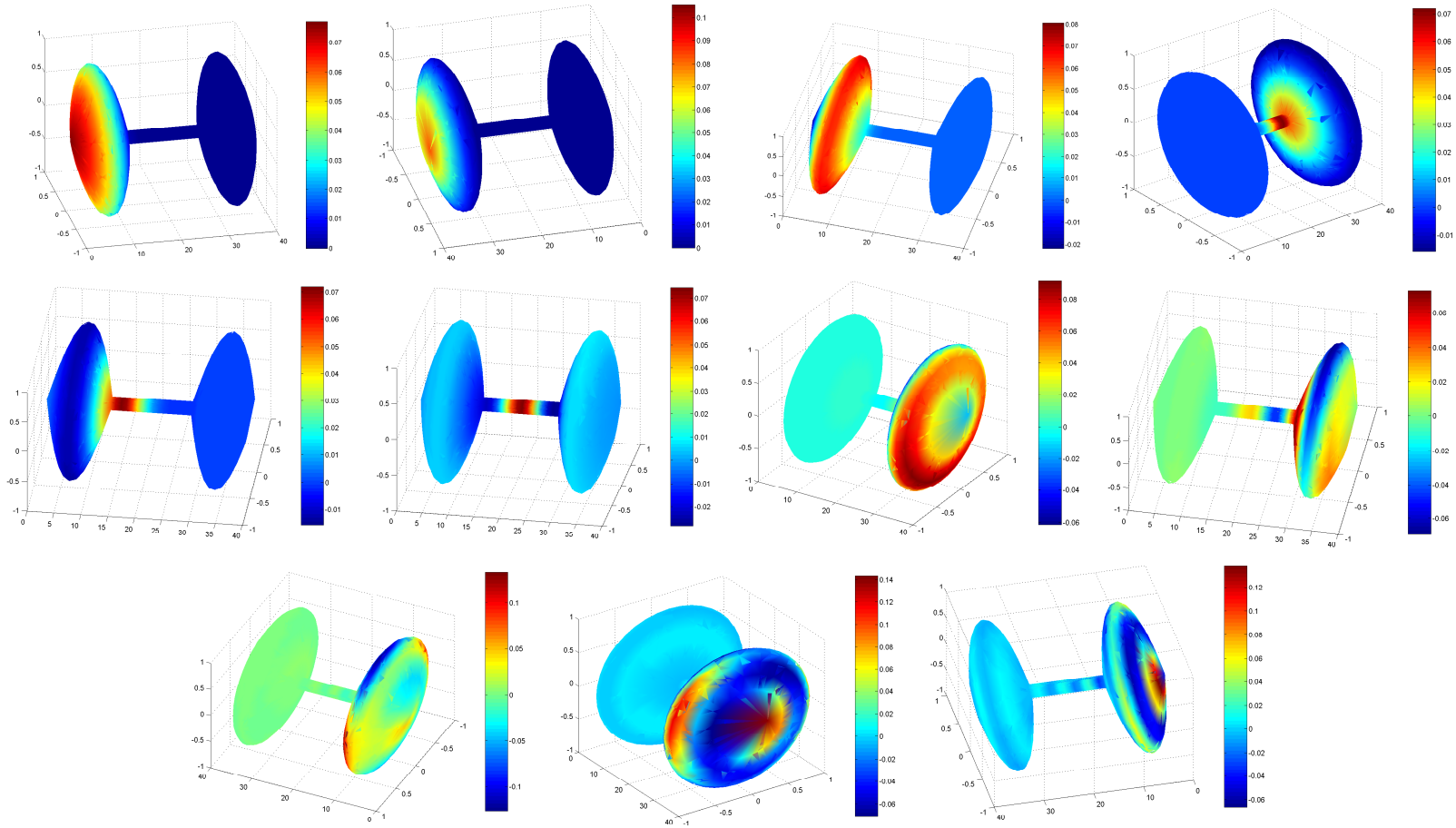
[Joint with RR Coifman]

Suppose for simplicity we have a weighted graph (G, E, W), with corresponding Laplacian $\mathcal{L}$ and random walk P. Let us renormalize, if necessary, P so it has norm 1 as an operator on L^2: let T be this operator. Assume for simplicity that T is self-adjoint, and high powers of T are low-rank: T is a diffusion, so range of T^t is spanned by smooth functions of increasingly (in t) smaller gradient.

A "typical" spectrum for the powers of T would look like this:



Diffusion Wavelets on Dumbell manifold

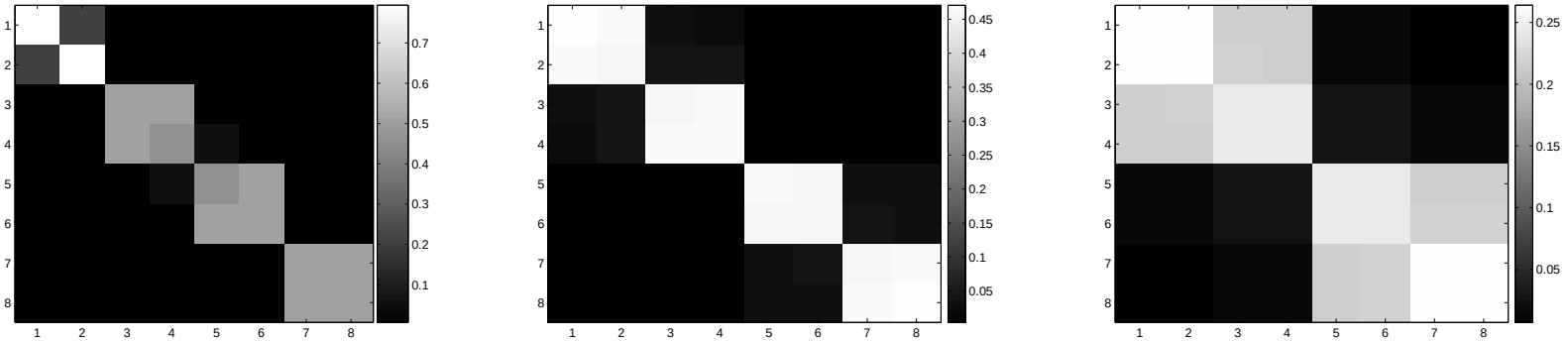


Coarsening of Markov chains

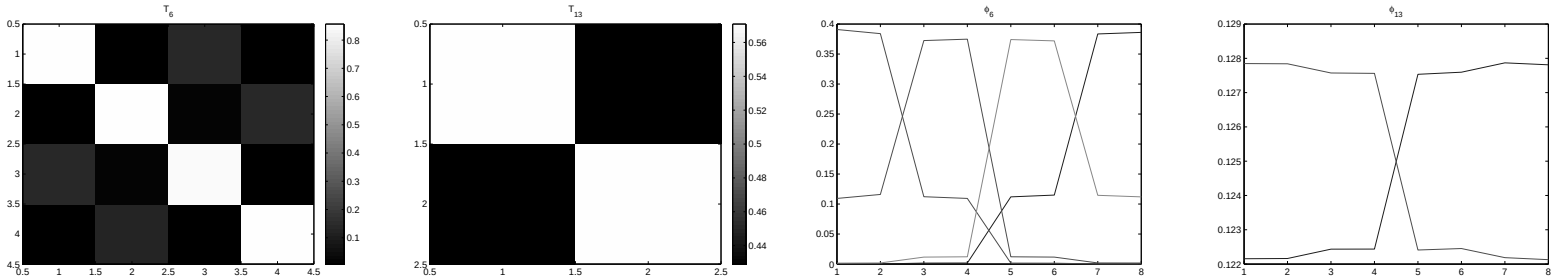
We now consider a simple example of a Markov chain on a graph with 8 states.

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

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.

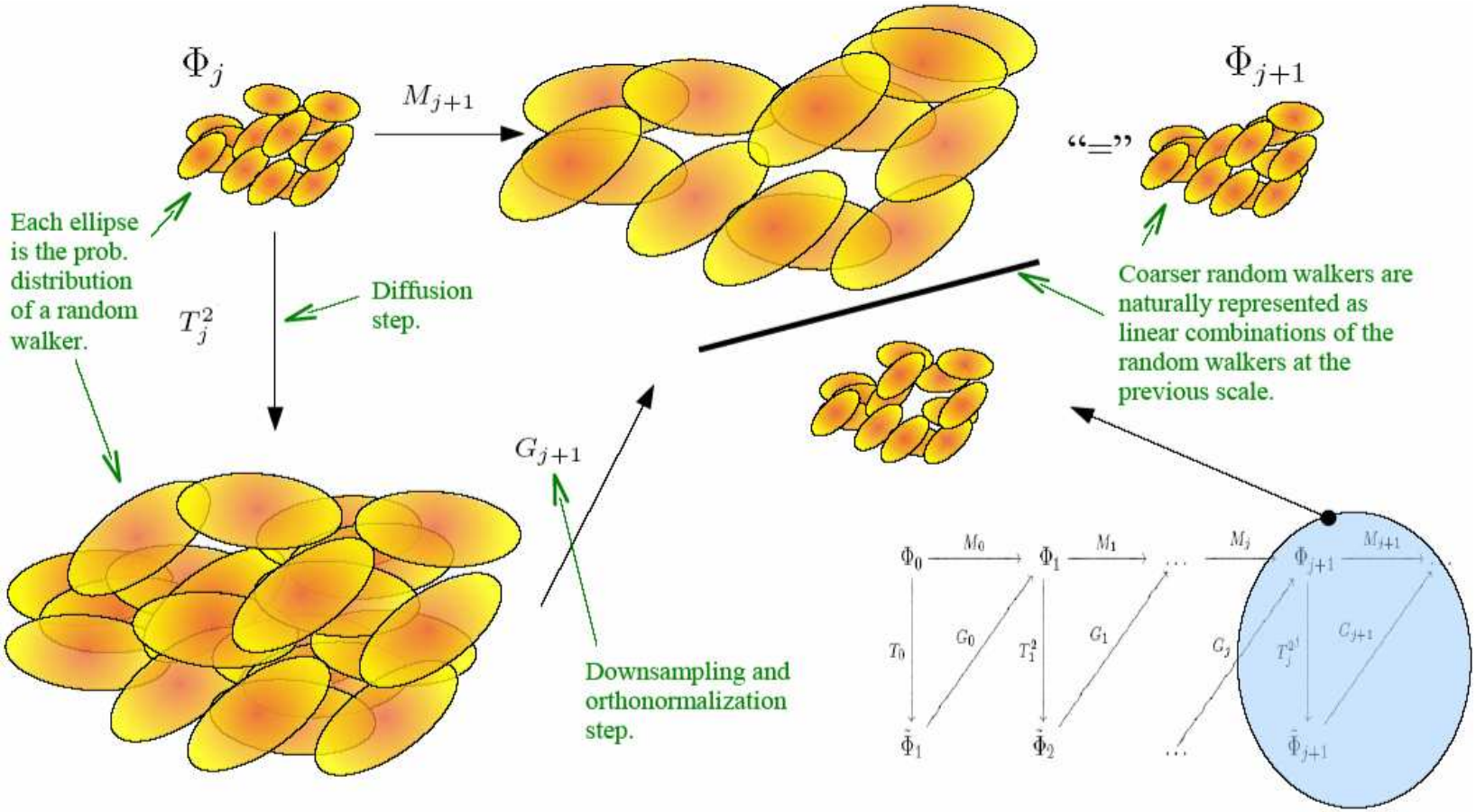


Some dyadic powers of the Markov chain T.



Compressed representations T_6, T_{13}, and corresponding scaling functions.

Construction of Diffusion Wavelets



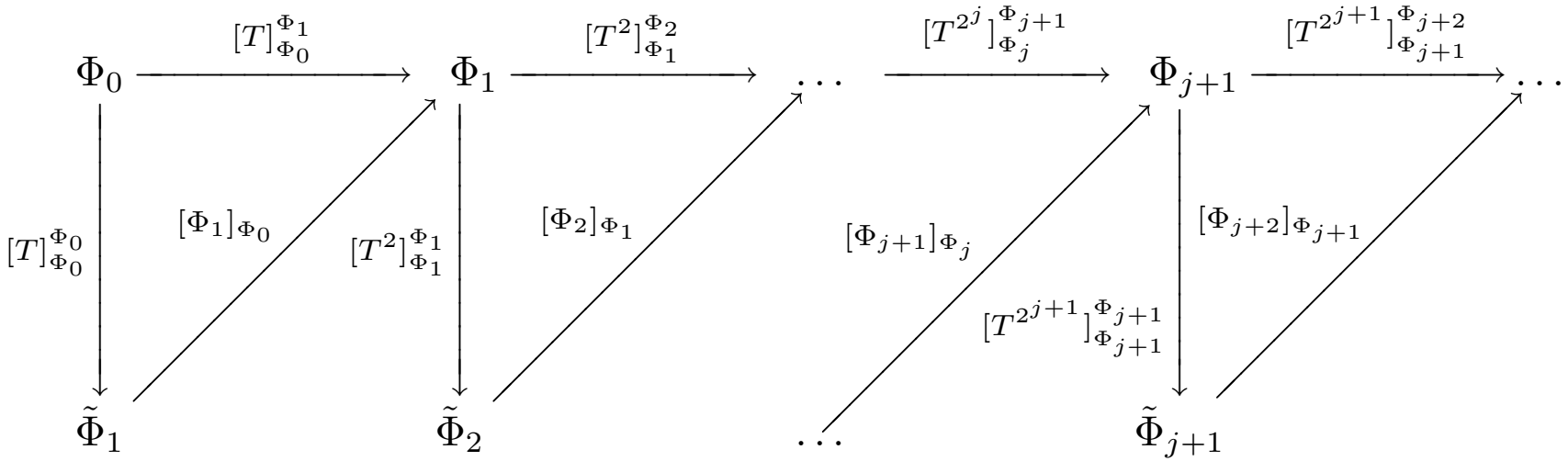


Diagram for downsampling, orthogonalization and operator compression. (All triangles are commutative by construction)

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

// $[T]_{\Phi_0}^{\Phi_0}$: 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, template below.

// ϵ: precision

// Output: The orthonormal bases of scaling functions, Φ_j, wavelets, Ψ_j, and

// compressed representation of T^{2^j} on Φ_j, for j in the requested range.

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

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

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

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

end

$Q, R \leftarrow$ `SpQR` (A, ϵ) // A: sparse $n \times n$ matrix, ϵ: precision

// Output: Q, R matrices, hopefully 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.

Thinking multiscale on graphs...

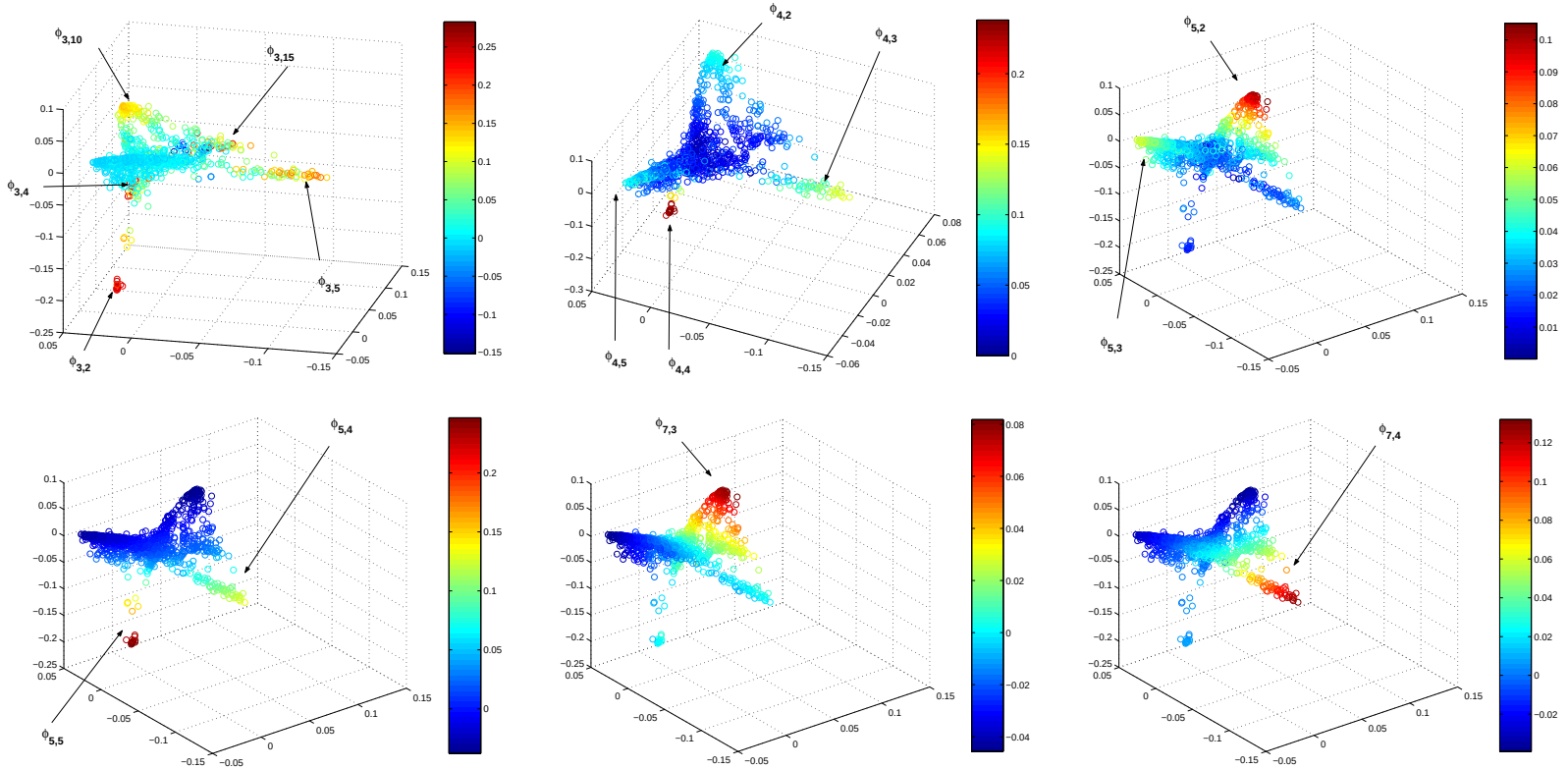
Investigating other constructions:

- Biorthogonal diffusion wavelets, in which scaling functions are probability densities (useful for multiscale Markov chains)
- Top-bottom constructions: recursive subdivision
- Both...

Applications:

- Document organization and classification
- Markov Decision Processes
- Nonlinear Analysis of Images
- Semi-supervised learning through diffusion processes on data

Application to text document Classification



Scaling functions at different scales represented on the set embedded in 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.

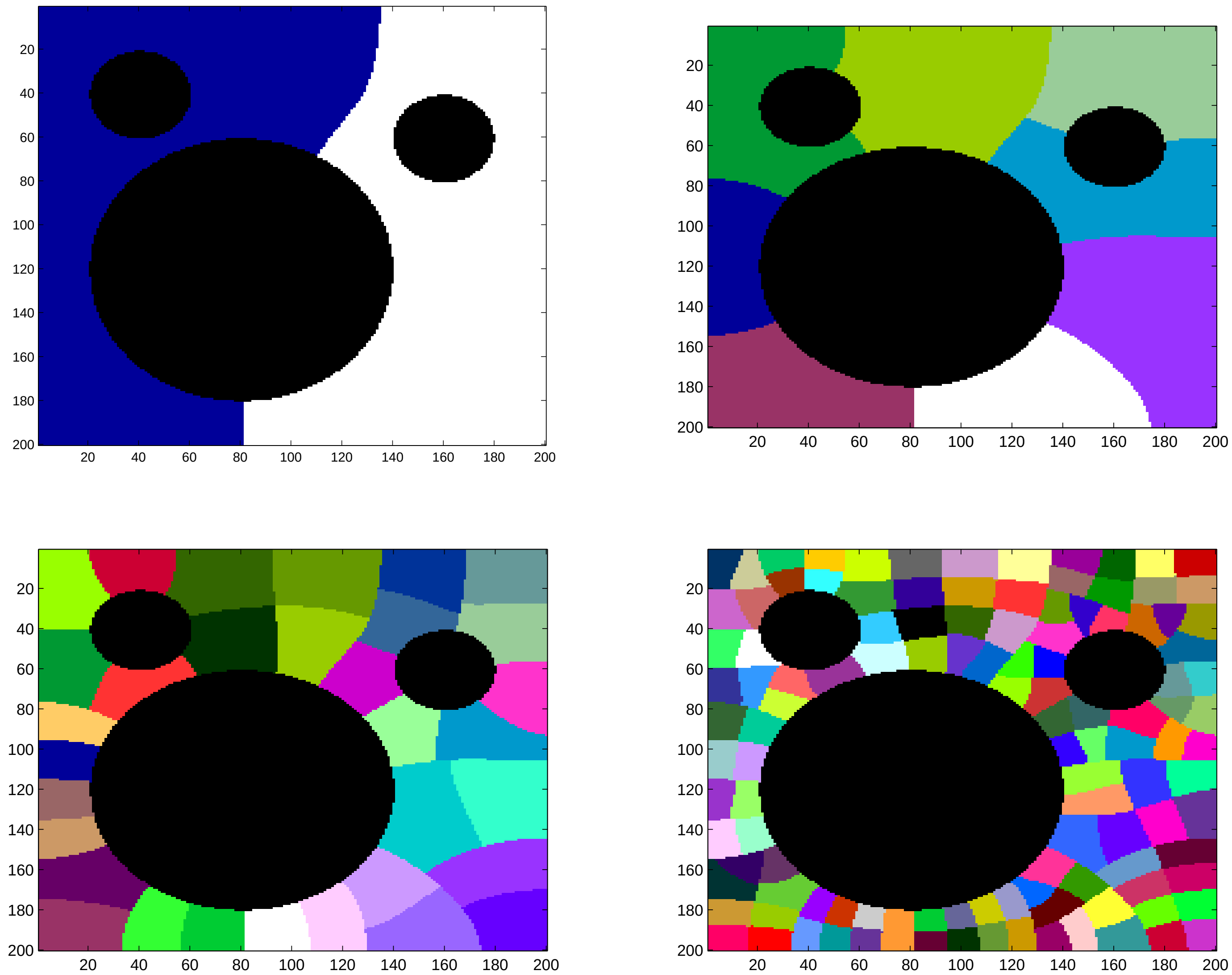
Top-bottom constructions

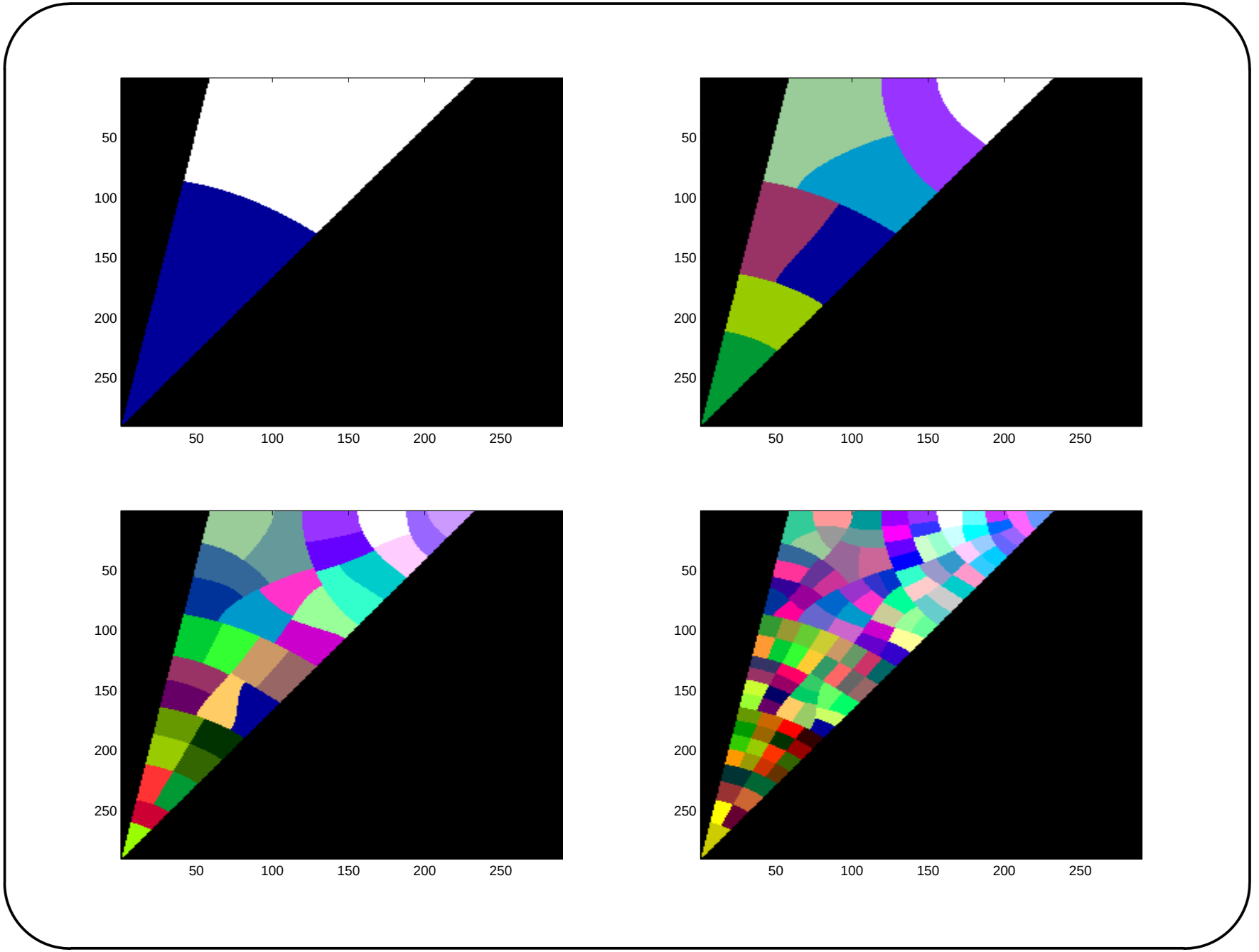
Spatial partitioning: use top (non-trivial) Neumann eigenfunction to cut the domain in two pieces, repeat on sub-domains, always by re-imposing Neumann boundary conditions. The Neumann nodal line is the minimizer of the asymptotic probability of escape of Brownian motion from a sub-domain, i.e. the subdivision of D into domains D_1 and D_2 that minimizes λ where there is a constants C_x so that

$$\lim_{t\to\infty} e^{-\lambda t} P(B_x \notin D_x) = C_x,$$

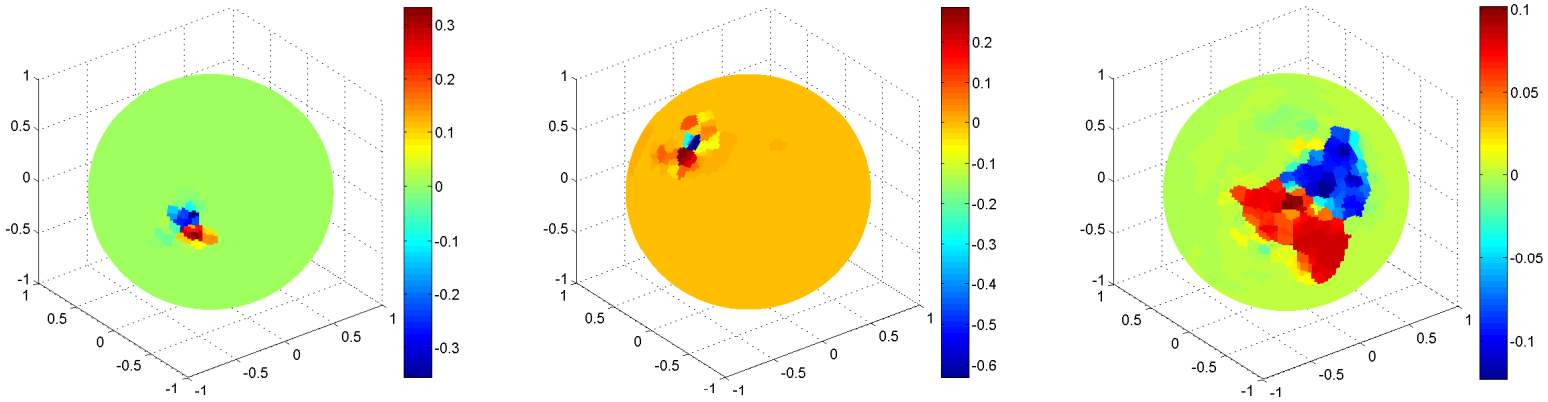
and where $x \in D_x$, B_x is Brownian motion started at x, reflected on ∂D, and killed on $\Gamma = \partial D_i - \partial D$. Thus the division according to the eigenfunction is like maximizing the volume to boundary area ratio as seen by a Brownian particle.

Gives binary tree, in practice, and at least in decent Euclidean domains, the pieces thus obtained are quite regular. At each scale we have a partition of the space into "dyadic" diffusion cubes.


20
40
60
80
100
120
140
160
180
200


50
100
150
200
250
50
100
150
200
250
50
100
150
200
250
50
100
150
200
250
50
100
150
200
250
50
100
150
200
250
50
100
150
200
250
50
100
150
200
250
50
100
150
200
250

Smoothed Haar functions



Three smoothed Haar functions on a sphere sampled randomly at 2000 points.

Application to Markov Decision Processes

[S. Mahadevan, MM]

A finite Markov decision process (MDP) $M = (S, A, P^a_{ss'}, R^a_{ss'})$ is defined as a finite set of states S, a finite set of actions A, a transition model $P^a_{ss'}$ specifying the distribution over future states s' when an action a is performed in state s, and a corresponding reward model $R^a_{ss'}$ specifying a scalar cost or reward. A state value function is a mapping $S \to \mathcal{R}$ or equivalently a vector in $\mathcal{R}^{|S|}$. Given a policy $\pi : S \to A$ mapping states to actions, its corresponding value function V^π specifies the expected long-term discounted sum of rewards received by the agent in any given state s when actions are chosen using the policy. Any optimal policy π^* defines the same unique optimal value function V^* which satisfies the nonlinear constraints

$$V^*(s) = \max_a \sum_{s'} P^a_{ss'} \left(R^a_{ss'} + \gamma V^*(s')\right)$$

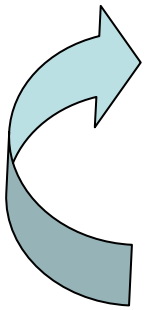
The state spaces of MDPs are often varifolds or graphs; it is crucial to represent certain functions and operators (large-time expectation operators $\sim$ Green's operators) efficiently.

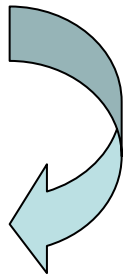
Policy Iteration

(Howard, PhD, MIT, 1959)

Policy Evaluation:
("Critic")

$$V^{\pi}(x) = r(x, \pi(x)) + \gamma \sum_{y} P_{xy}^{\pi(x)} V^{\pi}(y)$$

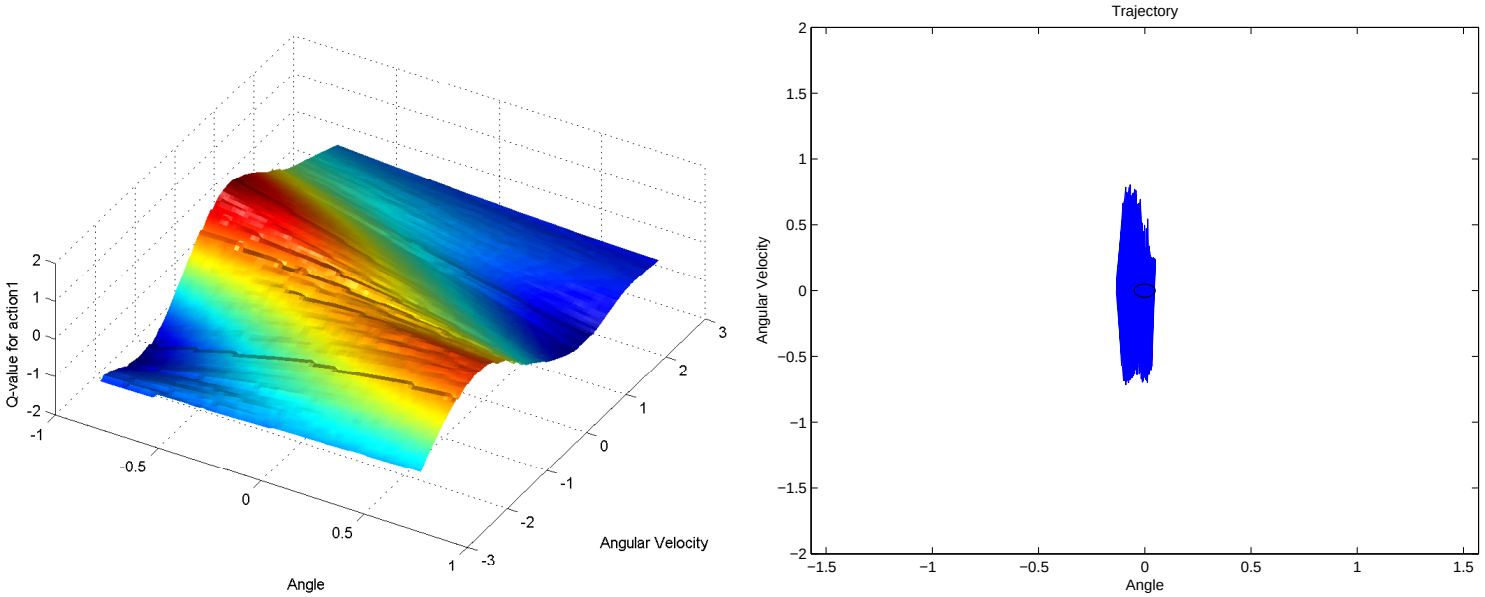




Policy Improvement:
("Actor")

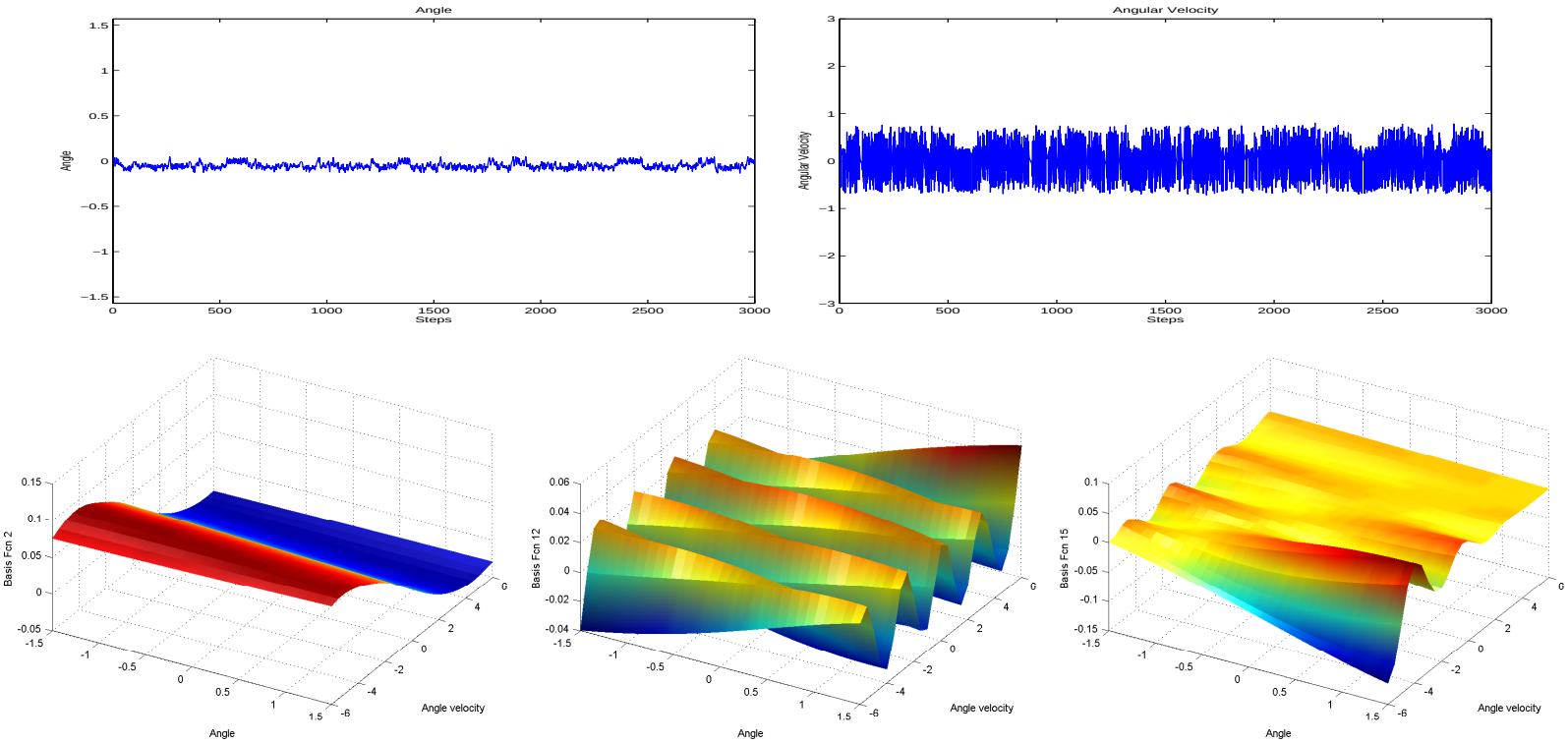
$$\pi'(x) = \operatorname{argmax}_a \left(r(x, a) + \gamma \sum_{y} P_{xy}^{a} V^{\pi}(y) \right)$$

Inverted Pendulum



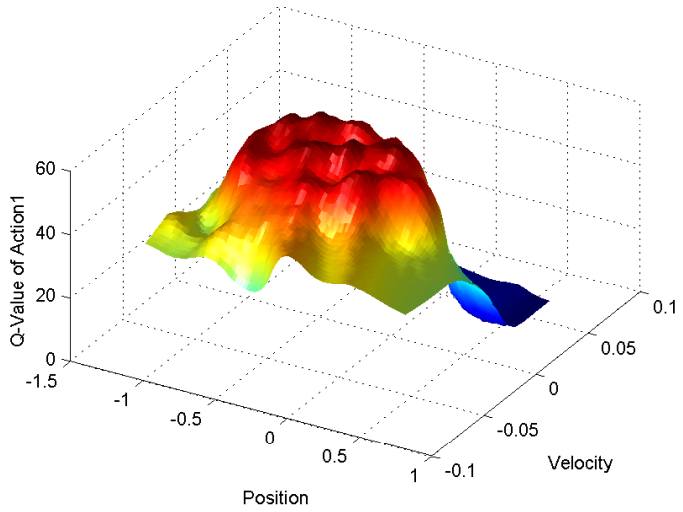
Left: Q-value function for the action "left", reconstructed from its representation of the diffusion wavelet basis. Right: trajectory of the pendulum in phase space according to the policy learnt.

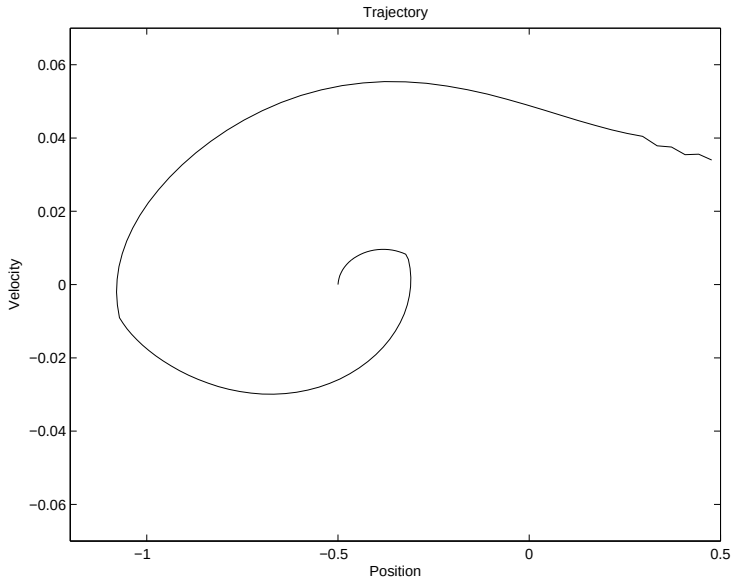
Inverted Pendulum (cont'd)

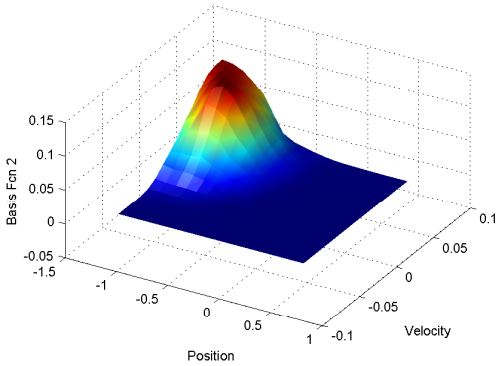


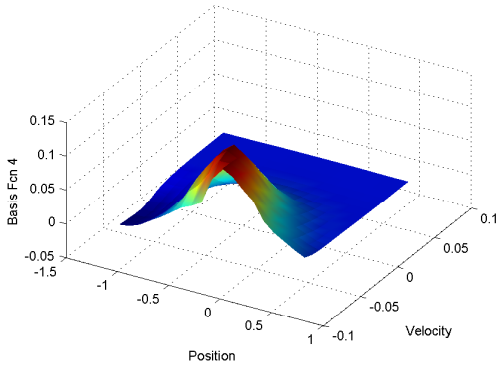
Top row: trajectory of angle and angle velocity variables. Bottom row: some diffusion wavelets used as basis functions for representation during the learning phase.

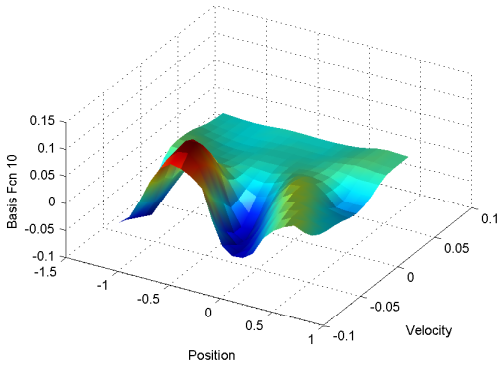
Mountain Car



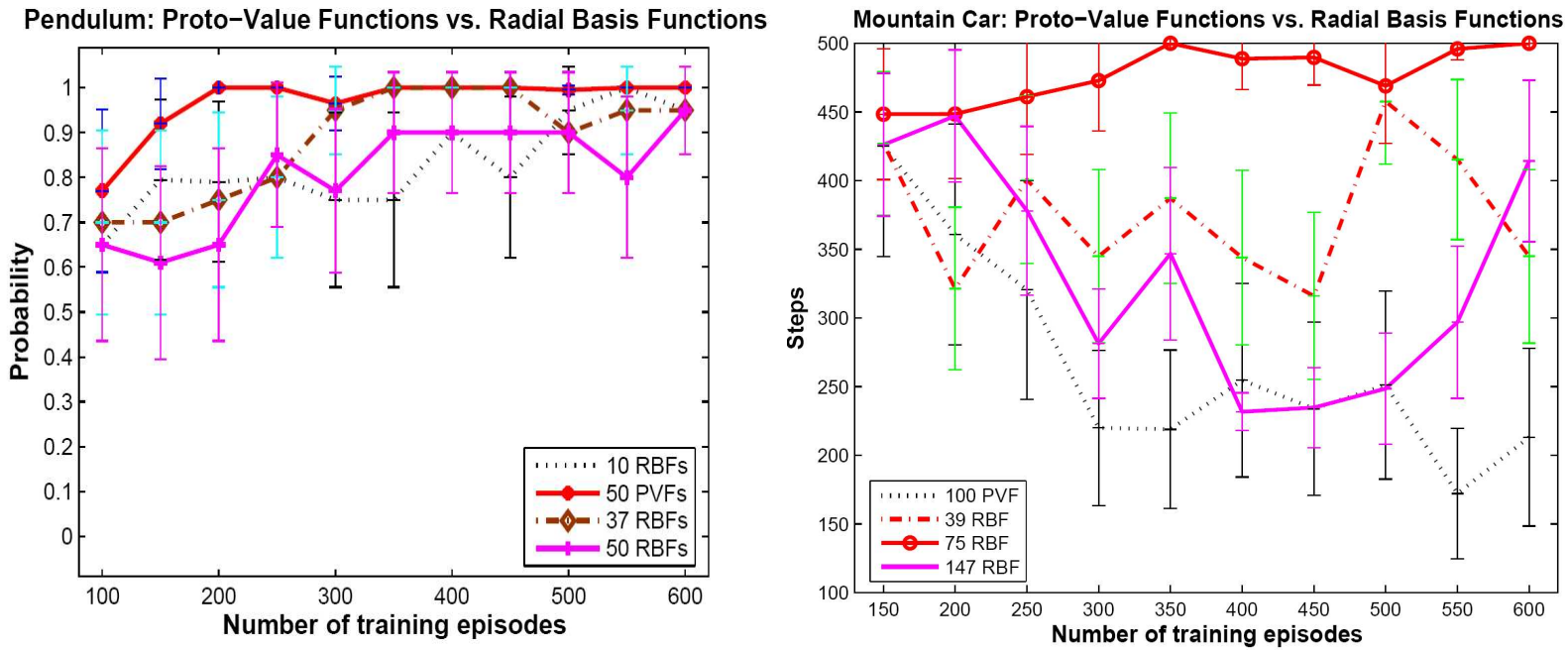








Top left: Q-value function for the action "left", reconstructed from its representation of the diffusion wavelet basis. Top right: trajectory of the mountain car in phase space according to the policy learnt (107 steps). Bottom row: some diffusion wavelets used as basis functions for representation during the learning phase.



Comparison between eigenfunctions and standard sets of basis functions for two environments: left an inverted pendulum, right the mountain car.

Semi-supervised Learning on Graphs

[Joint with A.D.Szlam]

Given a graph G and $\chi_1, \ldots, \chi_C$ characteristic functions of sets, representing points labeled according to their class (e.g. document topics, digit in a handwritten digit database, functionality of a protein in a protein network...). These labels are known on a small subset $\tilde{G}$ of G, and we would like to guess the labels of the non-labeled points. It is an interpolation or smoothing problem.

So far good results (in theory and in practice) with the use of eigenfunctions [Belkin,Niyogi], even better results by using anisotropic diffusions on the graph to smooth the label functions; just started work with diffusion wavelets for this task and other machine learning tasks.

We are applying these techniques to handwritten digits, documents classification, protein functionality prediction in protein networks.

Semi-supervised Learning on Graphs, cont'd

[Joint with A.D.Szlam]

Eigenfunctions [Belkin-Niyogi]: Predict unlabeled points by projecting onto a subspace spanned by low-frequency eigenfunctions, restricted to the labelled set $\tilde{G}$. Motivations & assumptions: the label functions are smooth w.r.t. geometry of space, eigenfunctions capture idea of smoothness with respect to the geometry.

New nonlinear technique: use diffusion process to smooth the label functions from $\tilde{G}$ to functions on G. Each point has now a vector of probabilities of belonging to different classes: use this extra information to design a better, anisotropic diffusion on G, and start anew by applying this to the initial labels. Motivations: the diffusion process is a much more flexible tool than eigenfunctions, for example it is easy to tune time-scales, it is easily tuned to incorporate labeling information, it has a better spectral properties than a spectral projector (e.g. no Gibbs phenomenon), moreover it is very fast to compute!

Experiments on USPS zip code data set show this technique outperforms the previous semi-supervised learning algorithms. We are applying this technique to a problem in prediction of *protein functionality*, where there are more than 40 classes, with encouraging preliminary results.

Current & Future work

Create a bag of tools for intrinsic analysis *of* and *on* data sets, mathematically modeled as rather general classes of possibly wild sets in Euclidean space, or graphs.

Ideally an analysis, tools, algorithms flexible enough to tackle different problems, from nonlinear dimensionality reduction to semisupervised and supervised learning and reinforcement learning.

Applications we are considering: signal processing on manifolds and graphs and its applications (e.g. linear and nonlinear "image" denoising); classification algorithms (e.g. text classification, protein and gene functional classification, target recognition in hyper-spectral imaging); learning; application to Markov decision processes; multiscale structures of complex networks and dynamical systems.

Collaborators

- R.R. Coifman, P.W. Jones (Yale Math) [Diffusion geometry; Diffusion wavelets; Uniformization via eigenfunctions; Multiscale Data Analysis], S.W. Zucker (Yale CS) [Diffusion geometry];
- 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];
- S. Mahadevan (U.Mass CS) [Markov decision processes];
- R. Schul (UCLA) [Uniformization via eigenfunctions; nonhomogenous Brownian motion];
- A.D. Szlam (Yale) [Diffusion wavelet packets, top-bottom multiscale analysis, linear and nonlinear image denoising, classification algorithms based on diffusion];
- Y. Kevrekidis (Princeton Eng.), S. Lafon (Google), B. Nadler (Weizman) [stochastic dynamics];
- W. Goetzmann (Yale, Harvard Business School), J. Walden (Berkley Business School), P.W. Jones (Yale Math) [Applications to finance]
- H. Mhaskar (Cal State, LA) [polynomial frames of diffusion wavelets, characterization of function spaces];
- A. Paccanaro and R. Sasidharan (Yale, MicroBio.) [Semi-supervised learning for protein network functionality prediction];
- J.C. Bremer (Yale) [Diffusion wavelet packets, biorthogonal diffusion wavelets];
- M. Mahoney (Yahoo Research), F. Meyer (UC Boulder), X. Shen (UC Boulder) [Randomized algorithms for hyper-spectral and fMRI imaging]

Material (Matlab code, tutorial talks) available at `www.math.yale.edu/~mmm82`.

Thank you!