Diffusion Wavelets for multiscale analysis on manifolds and graphs: constructions and applications

Mauro Maggioni

Nov. 7th, 2005
Overview

1 - Connections with classical Harmonic Analysis;
2 - Would like to generalize to point clouds, metric spaces, “data sets”: use heat kernel and diffusion;
3 - Fourier global analysis;
4 - Multiscale analysis.
Harmonic Analysis

1 - *Ambient Space*, such as $\mathbb{R}, \mathbb{R}^n$, Lie groups ($H^n,...$), hypergroups, manifolds (Lip. curves), ...

2 - *Function Spaces*, such as $L^p$, Sobolev spaces, Besov spaces, ...

3 - *Operators*, such as Fourier transform, $\Delta$, Hilbert transform, maximal operators, Cauchy integral, ...

4 - *Special bases for $L^2$*, such as Fourier basis, Gabor, wavelets, wave packets, ...

(4) allows to study (3) via (2). The (measure-geometric, algebraic...) properties (1) affect (2),(3),(4). Basic idea: matrix representing an operator is almost diagonal for good choices of (4).
Fast Algorithms

The ideas developed in Harmonic Analysis have had huge impact on the design of fast algorithms for the solution of integral and differential equations, geometric analysis, and more.

These algorithms take advantage of the new structures discovered in the harmonic analysis of these problems. Many of these structures are multiscale.

Example of such algorithms are fast multipole methods [Greengard, Rohklin,...] for the solution of certain integral equations from mathematical physics (e.g. Dirichlet problem, wave equation, Helmholtz equation) in domains of low-dimensional Euclidean space, multi-grid methods [Brandt, Hackbusch,...], non-standard wavelet representation for certain integral and pseudo-differential operators [Beylkin, Coifman, Rohklin], the wavelet-element methods [Dahmen, DeVore, Cohen,...], and many others.
Ambient Space

Classically one studies Euclidean spaces $\mathbb{R}, \mathbb{R}^n$ [Stein, Weiss].

From there, we may recognize three directions towards more general structures:

- algebraic/highly symmetric structures (Lie groups),
- smooth geometric structures (Riemannian manifolds), global analysis
- less smooth geometric structures (Lipschitz curves).

Harder, and “more fundamental”:

- classes of point sets in $\mathbb{R}^n$,
- classes of metric spaces, for example measure-metric-energy spaces or graphs.

New type of harmonic analysis needed.
New type of Harmonic Analysis needed

We would like to have function spaces (2), operators (3), bases (4) interacting as in classical Harmonic Analysis on more general structures such as point sets in $\mathbb{R}^n$ and metric spaces.

Classical definitions take advantage of symmetries, geometric transformations of the space (and associated representations), and other ingredients which are simply not available here. Need new definitions, new interactions between geometry, function spaces, operators and bases.

Additional difficulties: incorporate noise into the picture.
An example from Molecular Dynamics...

The dynamics of a small protein in a bath of water molecules is approximated by a Fokker-Planck 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, this set of points is much lower-dimensional. We are able to discover this lower-dimensional set, 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!
The Heat Kernel and the Laplacian

The heat kernel and the Laplacian are present almost everywhere in classical Harmonic Analysis. In Euclidean space the Laplacian is very natural because of its invariance under the natural symmetries of the space. The associated heat kernel is of great importance at the very least because of its connections with Brownian motion, potential theory and the heat equation.

These objects can be defined in a natural way also in much general settings, such as metric-measure-energy spaces and graphs. On graphs the Laplacian is in fact quite a natural object, in the sense that given the weights, it satisfies They are affected by geometric properties of the space they are defined on, and can be used (for example through its spectral theory) to define function spaces, operators, and bases that are natural generalization of their classical counterparts.
Laplacian on manifolds I

The Laplace-Beltrami operator $\Delta_{BL}$ can be defined naturally on a Riemannian manifold, and is a well-studied object in global analysis. The corresponding heat kernel $e^{-t\Delta}$ is the Green’s function for the heat equation on the manifold, associated with Brownian motion “restricted” to the manifold. Spectral decomposition $\Delta \phi_i = \lambda_i \phi_i$ yields

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

The eigenfunctions $\phi_i$ of the Laplacian generalize Fourier modes: Fourier analysis on manifolds, global analysis.

Surprisingly, they 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

$$\Phi^{(t)}_m(x) = (\mu_i^t \phi_i(x))_{i=1,...,m} \in \mathbb{R}^n.$$
Diffusion Maps and Diffusion Distance

The map is an approximate isometry (it is an isometry for \( m = +\infty \)) to Euclidean \( \mathbb{R}^m \) from \( \mathcal{M} \) with the diffusion metric (\textit{not} the Riemannian metric!) defined by

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

We can also prove that eigenfunctions provide local coordinate systems [P.W. Jones, R. Schul, MM].
Diffusion vs. Geodesic Distance

\[ d_{\text{geod.}}(A, B) \sim d_{\text{geod.}}(C, B), \text{ however } d^{(t)}(A, B) \gg d^{(t)}(C, B). \]
Diffusion maps: Example

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

Graph whose vertices are mathematicians, edge between two mathematicians if co-authored of a paper. The Erdős number of a mathematician M is the length of a geodesic path between M and the vertex corresponding to P. Erdős. The Erdős connected component has small diameter (13), the mean distance is < 5 and the volume is concentrated in a ball of radius 8 around Erdős.

This is not very informative; it is called the small world effect. One may try to argue that in some sense the world of mathematicians is small...however: \( \text{diam(} \text{world population} \text{)} \sim 20 \), \( \text{diam(} \text{www} \text{)} \sim 19 \). The diffusion distance will take into account not only the shortest path from Erdős to a mathematician, but all the other paths, of all lengths, and takes a weighted average of them, refining the classifications of connections.
Laplacian on Graphs, I

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 \(\lambda_i\) and \(\phi_i\) 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 \(D^{-1}W\).

If

- points are drawn from a manifold according to some (unknown) probability distribution [M. Belkin, P. Niyogi; RR Coifman, S. Lafon], or
- points are generated by a stochastic dynamical system driven by a Fokker-Planck equation [RR Coifman, Y. Kevrekidis, S. Lafon, MM, B. Nadler]

the Laplace-Beltrami operator and, respectively the Fokker-Planck operator, can be approximated by a graph Laplacian on the discrete set of points with certain appropriately estimated weights.
Example from Molecular Dynamics revisited

The dynamics of a small protein in a bath of water molecules is approximated by a Fokker-Planck 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.
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.
Weights from a local similarity kernel

The similarity between points of a set $X$ can be summarized in a kernel $K(x, y)$ on $X \times X$. Usually we assume the following properties of $K$:

$$K(x, y) = K(y, x) \quad \text{(symmetry)}$$

$$K(x, y) \geq 0 \quad \text{(positivity preserving)}$$

$$\langle v, Kv \rangle \geq 0 \quad \text{(positive semi-definite)}$$

If $X \subseteq \mathbb{R}^n$, then choices for $K$ include $e^{-\frac{||x-y||^2}{\delta}}$, $\frac{\delta}{\delta + ||x-y||}$, $\frac{\langle x, y \rangle}{||x|| ||y||}$.

If some “model” for $X$ is available, the kernel can be designed to be consistent with that model.

In several applications, one starts by applying a map to $X$ (projections onto lower-dimensional subspaces, nonlinear maps, etc...) before constructing the kernel.
A simple example

$X$ is a set of images of a symbol “3D” under various lighting conditions. Each point is a vector in $\mathbb{R}^{32^2}$, the $i$-th coordinate being the intensity of the $i$-th pixel in the image. We use the kernel $e^{-\left(|\frac{x-y}{\delta}|\right)^2}$, for a reasonable choice of $\delta$, which restricted to $X$ yields the weights on the graph of points. We compute the graph Laplacian and its low-frequency eigenfunctions, and use them to embed the data set in two-dimensions. The eigenfunctions actually discover the natural parametrization by the lighting vector.

Picture courtesy of Stephane Lafon
Several Applications

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

- Classifiers in the semi-supervised learning context [M. Belkin, P. Nyogi]
- 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 Hyper-spectral Pathology

For each pixel in a hyper-spectral image we have a whole spectrum (a 128-dimensional vector for example). We view the ensemble of all spectra in a hyper-spectral image as a cloud in $\mathbb{R}^{128}$, induce a Laplacian on the point set and use the eigenfunctions for classification of spectra into different classes, which turn out to be biologically distinct and relevant.

On the left, we have mapped the values of the top 3 eigenfunctions to RGB.
Application to text document classification

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 $\Xi^{(0)}_6(x) = (\xi_1(x), \ldots, \xi_6(x))$: 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!
Multiscale Analysis, I

We would like to construct multiscale bases, generalizing classical wavelets, on manifolds, graphs, point clouds.

The classical construction is 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.
Suppose for simplicity we have a weighted graph \((G, E, W)\), with corresponding Laplacian \(L\) 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:
Classical Multi-Resolution Analysis

A Multi-Resolution Analysis for $L^2(\mathbb{R})$ is a sequence of subspaces $\{V_j\}_{j \in \mathbb{Z}}$, $V_{j-1} \subseteq V_j$, with $\cap V_j = \{0\}$, $\overline{\cup V_j} = L^2(\mathbb{R})$, with an orthonormal basis $\{\phi_{j,k}\}_{k \in \mathbb{Z}} := \{2^{\frac{j}{2}} \phi(2^j \cdot -k)\}_{k \in \mathbb{Z}}$ for $V_j$. Then there exist $\psi$ such that $\{\psi_{j,k}\}_{k \in \mathbb{Z}} := \{2^{\frac{j}{2}} \psi(2^j \cdot -k)\}_{k \in \mathbb{Z}}$ spans $W_j$, the orthogonal complement of $V_{j-1}$ in $V_j$.

$\hat{\phi}_{j,k}$ is essentially supported in $\{|\xi| \leq 2^j\}$, and $\hat{\psi}_{j,k}$ is essentially supported in the L.P.-annulus $2^{j-1} \leq |\xi| \leq 2^j$.

Because $V_{j-1} \subseteq V_j$, $\varphi_{j-1,0} = \sum_{k'} \alpha_{k'} \varphi_{j,k'}$: refinement eqn.s, FWT.

We would like to generalize this construction to graphs. The frequency domain is the spectrum of $e^{-\mathcal{L}}$. Let $V_j := \langle \{\phi_i : \lambda_i^{2^j} \geq \epsilon\} \rangle$. Would like o.n. basis of well-localized functions for $V_j$, and to derive refinement equations and downsampling rules in this context.
Construction of Diffusion Wavelets

Each ellipse is the prob. distribution of a random walker.

Diffusion step.

Downsampling and orthonormalization step.

Coarser random walkers are naturally represented as linear combinations of the random walkers at the previous scale.
Some diffusion wavelets and wavelet packets on the sphere, sampled randomly uniformly at 2000 points.
Diffusion Wavelets on Dumbell manifold
Signal Processing on Manifolds

Left: reconstruction of the function $F$ with top 50 best basis packets. Right: 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.
Potential Theory, Efficient Direct Solvers

The Laplacian $\mathcal{L} = I - T$ has an inverse (on $\ker(\mathcal{L})^\perp$) whose kernel is the Green’s function, that if known would allow the solution of the Dirichlet or Neumann problem (depending on the boundary conditions imposed on the problem on $\mathcal{L}$). If $||T|| < 1$, one can write the Neumann series

$$(I - T)^{-1} f = \sum_{k=1}^{\infty} T^k f = \prod_{k=0}^{\infty} (I + T^{2^k}) f.$$  

Since we have compressed all the dyadic powers $T^{2^k}$, we have also computed the Green’s operator in compressed form, in the sense that the product above can be applied directly to any function $f$ (or, rather, its diffusion wavelet transform). Hence this is a direct solver, and potentially offers great advantages, especially for computations with high precision, over iterative solvers.
Summary of the Algorithm

Input: A diffusion operator represented on some orthonormal basis (e.g.: \( \delta \)-functions), and a precision \( \epsilon \).

Output: Multiscale orthonormal scaling function bases \( \Phi_j \) and wavelet bases \( \Psi_j \), encoded through the corresponding multiscale filters \( M_j \), as well as \( T^{2^j} \) represented (compressed) on \( \Phi_j \).

One can prove that \( \Psi_j \) is like a Littlewood-Paley block in frequency. Currently working out the implications on characterization of function spaces and corresponding approximation properties.

Can also construct wavelet packets for a more flexible space-frequency analysis.

Allows for a fast wavelet transform, best basis algorithms, signal processing on graphs and manifolds, efficient application of \( T^{2^j} \), and direct inversion of the Laplacian (next slide).
Application to Document Classification, revisited

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.
Application to Markov Decision Processes

The state spaces of MDPs are often varifolds or graphs; it is crucial to represent certain functions (value function=potential of certain rewards) and operators (large-time expectation operators \( \sim \) Green’s operators) efficiently. Excellent results obtained so far [S. Mahadevan, MM] with eigenfunctions of certain Laplacians and diffusion wavelets.

Eigenfunctions of the Laplacian in a 3 room environment (left), diffusion wavelets in a 2-room environment (right).
Current & Future work

Understand the complex relationships between the geometric measure theory aspects of point clouds in high dimensions and related multiscale geometric algorithms, the construction (or estimation) of the heat kernel, the multiscale functional construction of diffusion wavelets, and potential theory aspects.

Small steps at a time: I am working on better constructions of diffusion wavelets (from the computational perspective); top-bottom constructions; robustness of these constructions to noise; approximation theory and characterization of function spaces through diffusion wavelets.

Applications: 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.
Collaborators

- R.R. Coifman, P.W. Jones (Yale Math) [Diffusion geometry; Diffusion wavelets; Uniformization via eigenfunctions], S.W. Zucker (Yale CS) [Diffusion geometry];
- G.L. Davis, 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];
- Y. Kevrekidis (Princeton Eng.), S. Lafon (Google), B. Nadler (Weizman) [stochastic dynamics];
- A.D. Szlam (Yale) [Diffusion wavelet packets, top-bottom multiscale analysis, linear and nonlinear image denoising, classification algorithms based on diffusion];
- J.C. Bremer (Yale) [Diffusion wavelet packets, biorthogonal diffusion wavelets];
- R. Schul (UCLA) [Uniformization via eigenfunctions; nonhomogenous Brownian motion];
- H. Mashkar (LA State) [polynomial frames of diffusion wavelets, characterization of function spaces];
- M. Mahoney (Yahoo Research), F. Meyer (UC Boulder), X. Shen (UC Boulder) [Randomized algorithms for hyper-spectral and fMRI imaging]
- W. Goetzmann (Yale, Harvard Business School), J. Walden (Berkley Business School), P.W. Jones (Yale Math) [Applications to finance]

This talk, papers, Matlab code (currently working on Matlab toolbox) available at

www.math.yale.edu/~mmm82

Thank you!