Tutorial - Random Walks on Graphs
Multiscale Aspects

Mauro Maggioni

Mathematics and Computer Science
Duke University

I.P.A.M., 09/11/08

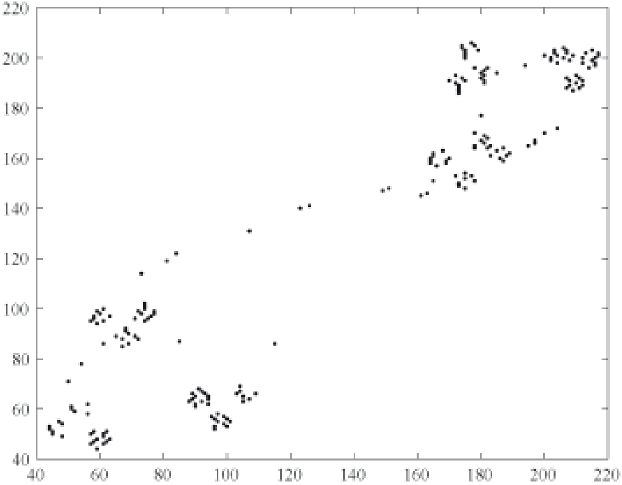
In collaboration with R.R. Coifman, P.W. Jones, Y-M. Jung, R. Schul, A.D. Szlam

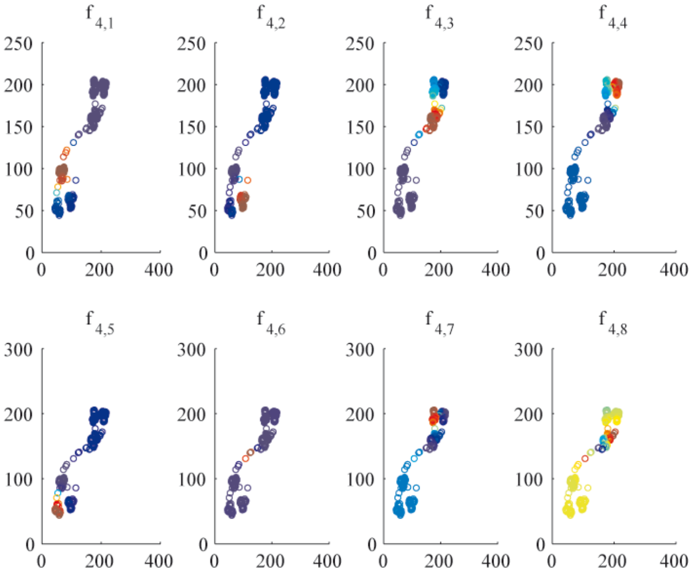
Funding: ONR, NSF

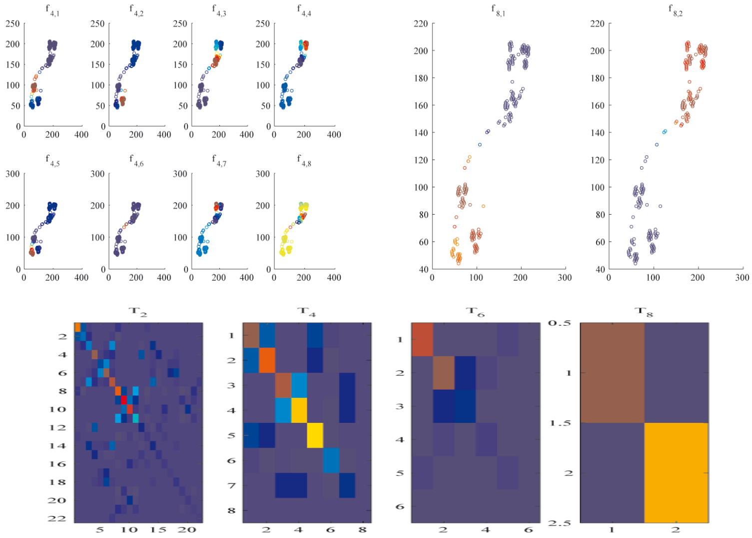
Plan

- Multiscale analysis
  - Multiscale construction
  - Geometric and functional interpretation
  - Diffusion wavelets and algorithms
- Examples
- Conclusion

A multiscale "network"




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



Multiscale elements and representation of powers of T

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.

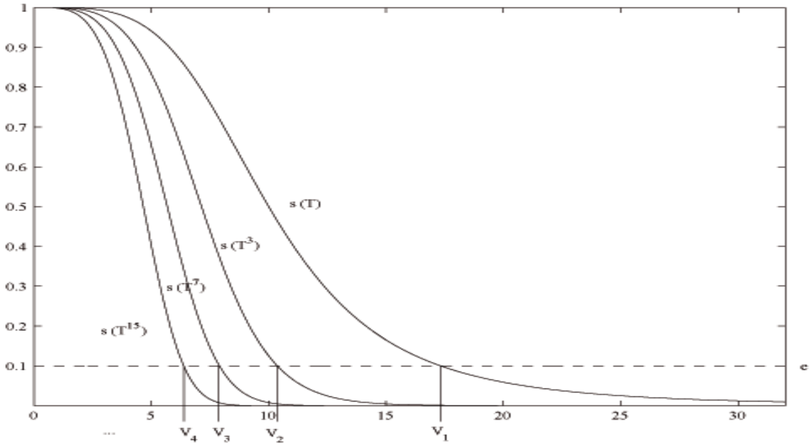
Multiscale Analysis, a bit more precisely

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.

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:



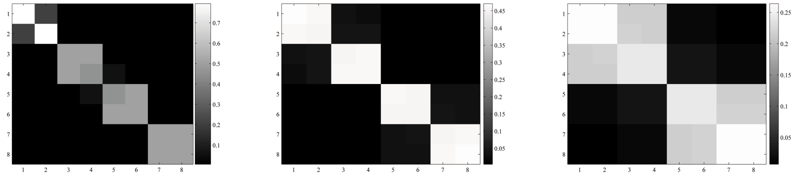
Multiscale Analysis, a trivial example, I

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

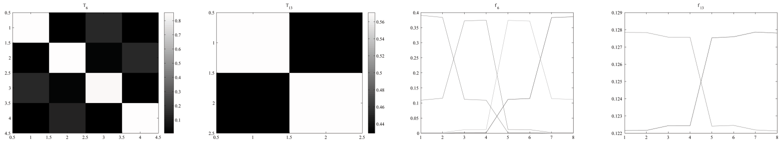
$$T = \begin{pmatrix} 0.80 & 0.20 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.20 & 0.79 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.49 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.50 & 0.499 & 0.001 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.001 & 0.499 & 0.50 & 0.00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

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

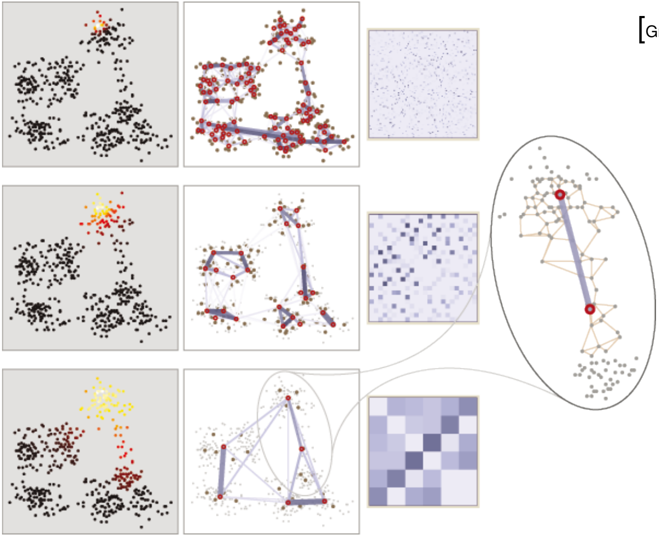


Some powers of the Markov chain T, 8×8, of decreasing effective rank.



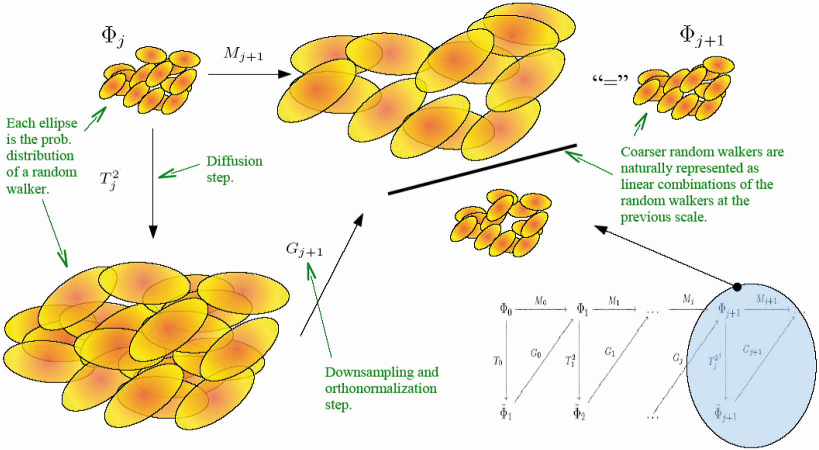
Compressed representations $T_6 := T^{2^6}$ (4×4), $T_{13} := T^{2^{13}}$ (2×2), and corresponding soft clusters.

Multiscale Analysis, a sketch



[Graphics by E. Monson]

Construction of Diffusion Wavelets, I



Construction of Diffusion Wavelets, II

$$\begin{tikzcd}
\Phi_0 \arrow[r, "{[T]_{\Phi_0}^{\Phi_1}}"] \arrow[d, "{[T]_{\Phi_0}^{\Phi_0}}"'] & \Phi_1 \arrow[r, "{[T^2]_{\Phi_1}^{\Phi_2}}"] \arrow[d, "{[T^2]_{\Phi_1}^{\Phi_1}}"'] & \ldots \arrow[r, "{[T^{2^j}]_{\Phi_j}^{\Phi_{j+1}}}"] & \Phi_{j+1} \arrow[r, "{[T^{2^{j+1}}]_{\Phi_{j+1}}^{\Phi_{j+2}}}"] \arrow[d, "{[T^{2^{j+1}}]_{\Phi_{j+1}}^{\Phi_{j+1}}}"'] & \ldots \\
\tilde{\Phi}_1 \arrow[ur, "{[\Phi_1]_{\Phi_0}}"] & \tilde{\Phi}_2 \arrow[ur, "{[\Phi_2]_{\Phi_1}}"] & \ldots \arrow[ur, "{[\Phi_{j+1}]_{\Phi_j}}"] & \tilde{\Phi}_{j+1} \arrow[ur, "{[\Phi_{j+2}]_{\Phi_{j+1}}}"] &
\end{tikzcd}$$

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

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

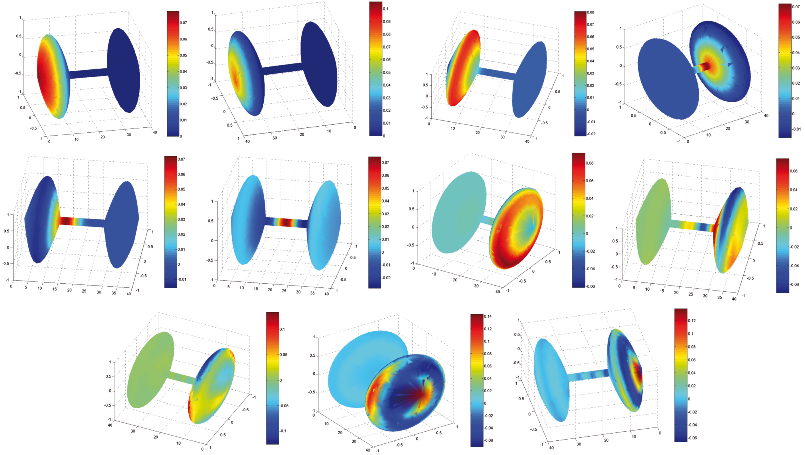
Signal processing tasks by adjusting wavelet coefficients.

Properties of Diffusion Wavelets

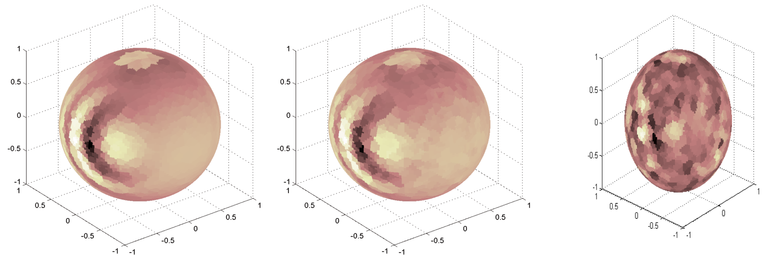
- Multiscale analysis and wavelet transform
- Compact support and estimates on support sizes (not as good as one really would like!);
- Vanishing moments (w.r.t. low-frequency eigenfunctions);
- Bounds on the sizes of the approximation spaces (depend on the spectrum of T, which in turn depends on geometry);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

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

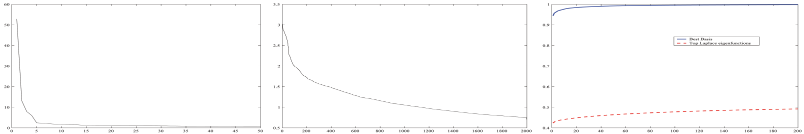
Diffusion Wavelets on Dumbell manifold



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.

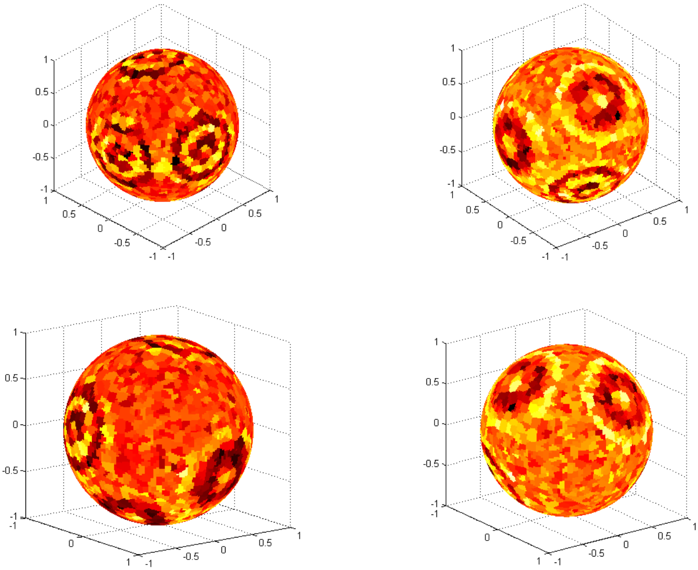


Left to right: 50 top coefficients of F in its best diffusion wavelet basis, distribution coefficients$F$ in the delta basis, first 200 coefficients of F in the best basis and in the basis of eigenfunctions.

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



Local Discriminant Bases, II

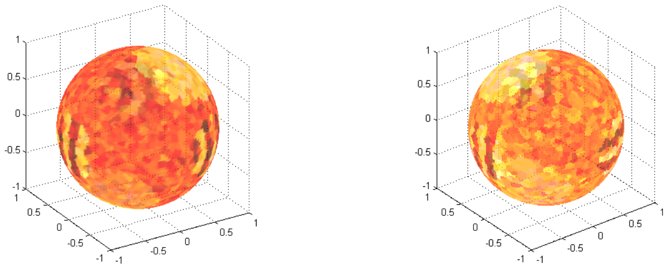
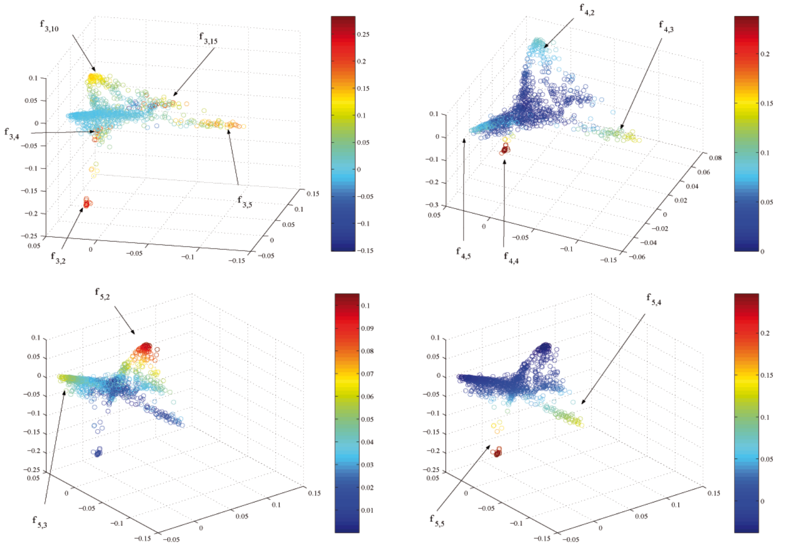


Figure: Left to right, a realization of a function from class 1 and 2 respectively. Note that the third smooth texture patch is on the back side of the sphere, and can be viewed in semitransparency. The other two smooth patches are decoys in random non-overlapping positions.

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.

Potential Theory, Compressed 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.

Many open questions and applications

- How do properties of diffusion wavelets relate to geometric (multiscale) properties of graphs?
- How to visualize these multiscale decompositions?
- Better constructions?

Applied to

- Multiscale signal processing (compression, denoising, discrimination) on graphs
- Multiscale learning on graphs
- Hierarchical clustering on nonlinear data sets.
- ...
- We will see at least a couple of applications to the analysis of networks and network traffic in other 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]; D. Brady, R. Willett (Duke EE) [Compressed sensing and imaging]

Funding: ONR, NSF.

Thank you!

`www.math.duke.edu/~mauro`