Harmonic and Multiscale Analysis of, and on, Data Sets

Mauro Maggioni
Dept. of Mathematics,
Computer Science & C.T.M.S.
Duke University

SAMPTA 2009, Luminy

Outline

Motivations & examples: large, high-dimensional data sets and graphs

Random walks, eigenfunctions, spectral embeddings

Multiscale analysis and diffusion wavelets

Applications
Structured data in high-D

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

Common feature: 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.
Function approximation on data

The geometry of the data may help the construction of useful priors, for prediction problems.

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.
Example 1: 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 dictionary.
Example 1: 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 dictionary.
Example 1: 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 dictionary.
Example 1: 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 dictionary.
Example 1: 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 dictionary.
Example 2: Handwritten digits

Data base of about 60,000 28x28 gray-scale pictures of handwritten digits, collected by USPS. Point cloud in $\mathbb{R}^{728}$. Goal: automatic recognition.
Example 2: Handwritten digits

Data base of about 60,000 28x28 gray-scale pictures of handwritten digits, collected by USPS. Point cloud in $\mathbb{R}^{728}$. Goal: automatic recognition.
Example 2: Handwritten digits

Data base of about 60,000 28x28 gray-scale pictures of handwritten digits, collected by USPS. Point cloud in $\mathbb{R}^{728}$. Goal: automatic recognition.
Example 2: Handwritten digits

Data base of about 60,000 28x28 gray-scale pictures of handwritten digits, collected by USPS. Point cloud in $\mathbb{R}^{728}$. Goal: automatic recognition.
Example 3: Molecular Dynamics

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 set of points in \( \mathbb{R}^{36} \)
Example 3: Molecular Dynamics

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 set of points in \( \mathbb{R}^{36} \)
Objectives

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 on data & graphs

Given:

- **Data** $X = \{x_i\}_{i=1}^N \subset \mathbb{R}^D$.

- **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

$$
P = D^{-1}W, \quad T = D^{-\frac{1}{2}} WD^{-\frac{1}{2}}, \quad H = e^{-tL}
$$

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.
Some properties of r.w.’s

- $P^t(x, y)$ 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 $1 = \lambda_1 \geq \lambda_2 \geq \ldots$.
- $\lambda_i \in [-1, 1]$
- “typically” $P$ (or $T$) is large and sparse, but its high powers are full and low-rank
Basic norms on graphs

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

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

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,

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

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

Several techniques rely on a mapping

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

where $\varphi_i$ 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 $T$: $T\varphi_i = \lambda_i \varphi_i$. 
Assume the data lies on a $d$-dimensional manifold $\mathcal{M}$ in $\mathbb{R}^n$ (think $n >> d$): how to find a map $\mathcal{M} \rightarrow \mathbb{R}^D$, with $D << n$ (hopefully $D \sim d$)?

Several techniques rely on a mapping

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

where $\varphi_i$ 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 $T$: $T\varphi_i = \lambda_i \varphi_i$.

Run some live examples!
Towards a theorem

In our situation, assuming the points are sampled from a manifold $\mathcal{M}$ of dim. $d$, as $n$, the number of points, grows, $L$ approaches $\Delta_{\mathcal{M}}$ (and $T$ approaches the heat kernel $K_{\mathcal{M}}$).

We are interested in maps $\mathcal{M} \to \mathbb{R}^d$ that are:

- **bi-Lipschitz** from $(\mathcal{M}, g)$ to $(\mathbb{R}^d, \| \cdot \|)$, so that distances are not distorted too much

- **defined on large pieces** of $\mathcal{M}$, so that we do not need too many charts to get an atlas of $\mathcal{M}$ (and local charts are easy to obtain anyway).
Eigenfcn’s Map
Eigenfcn's Map

$(\mathcal{M}, g)$

$\dim \mathcal{M} = d$
Eigenfcn’s Map

\[(M, g)\]

\[\dim M = d\]
Eigenfcn.'s Map

$$(\mathcal{M}, g) \quad \dim \mathcal{M} = d$$
Eigenfcn's Map

\((\mathcal{M}, g)\) 
\(\dim \mathcal{M} = d\)
Eigenfcn's Map

\[ \dim \mathcal{M} = d \]
Eigenfcn's Map

\( (\mathcal{M}, g) \)

\( \dim \mathcal{M} = d \)

\[ \Phi(x) = (\varphi_{i_1}(x), \ldots, \varphi_{i_k}(x)) \]
Eigenfcn's Map

\( (\mathcal{M}, g) \)
\( \dim \mathcal{M} = d \)

\[ \Phi(x) = (\varphi_{i_1}(x), \ldots, \varphi_{i_k}(x)) \]
Eigenfcn.'s Map

\[
(\mathcal{M}, g) \quad \text{dim} \mathcal{M} = d
\]

\[
\Phi(x) = (\varphi_{i_1}(x), \ldots, \varphi_{i_k}(x))
\]

Saturday, May 23, 2009
\[ \mathcal{M} \] smooth, \( d \)-dimensional compact connected manifold, possibly with boundary. Metric tensor \( g \in C^\alpha \). Fix \( z_0 \in \mathcal{M} \), let \( (U, F) \) be a coordinate chart s.t. \( z_0 \in U \) and normalized so that \( g^{il}(F(z_0)) = \delta^{il} \). Assume that for any \( x \in U, \xi, \nu \in \mathbb{R}^d \),

\[
\frac{c_{\min}(g)}{||\xi||_{\mathbb{R}^d}^2} \leq \sum_{i,j=1}^{d} g^{ij}(F(x))\xi_i \xi_j , \quad \sum_{i,j=1}^{d} g^{ij}(F(x))\xi_i \nu_j \leq \frac{c_{\max}(g)}{||\xi||_{\mathbb{R}^d}||\nu||_{\mathbb{R}^d}}
\]

Let \( r_U(z_0) = \sup\{r > 0 : B_r(F(z_0)) \subseteq F(U)\} \).

Recall Weyl’s estimate: let \( C_{\text{count}} \) be s.t. for any \( T > 0 \)

\[
\#\{j : 0 < \lambda_j \leq T\} \leq C_{\text{count}}T^{\frac{d}{2}}|\Omega|.
\]

[In the Dirichlet case \( C_{\text{count}} \) does not depend on \( \Omega \). Neumann case is more delicate.]
Theorem Let \((\mathcal{M}, g), \ z \in \mathcal{M}\) be a \(d\) dimensional manifold and \((U, F)\) be a chart as above. Assume \(|\mathcal{M}| = 1\). There is a constant \(\kappa > 1\), depending on \(d\), \(c_{\text{min}}, c_{\text{max}}, \|g\|_\alpha, \alpha\), such that the following hold.

Let \(\rho \leq r_U(z)\), then \(\exists i_1, \ldots, i_d:\) if \(\gamma_l = \left(\int_{B(z, \kappa^{-1}\rho)} \varphi_{i_l}^2 \right)^{-\frac{1}{2}}, \ l = 1, \ldots, d:\)

(i) the map \(\Phi : B(z, \kappa^{-1}\rho) \rightarrow \mathbb{R}^d\)

\[x \mapsto (\gamma_1 \varphi_{i_1}(x), \ldots, \gamma_d \varphi_{i_d}(x))\]

satisfies for any \(x_1, x_2 \in B(z, \kappa^{-1}\rho)\)

\[\frac{\kappa^{-1}}{\rho} d_{\mathcal{M}}(x_1, x_2) \leq \|\Phi(x_1) - \Phi(x_2)\| \leq \frac{\kappa}{\rho} d_{\mathcal{M}}(x_1, x_2).\]

(ii) \(\kappa^{-1}\rho^{-2} \leq \lambda_{i_1}, \ldots, \lambda_{i_d} \leq \kappa\rho^{-2}, \ \gamma_1, \ldots, \gamma_d \leq \kappa C_{\text{count}}^{\frac{1}{2}}.\)
Heat Kernel Map

$(\mathcal{M}, g)$

$\dim \mathcal{M} = d$
Heat Kernel Map

$$(M, g) \quad \dim M = d$$
Heat Kernel Map

$(\mathcal{M}, g)$

$\text{dim} \mathcal{M} = d$
Heat Kernel Map

$(\mathcal{M}, g)$

$\dim \mathcal{M} = d$

$R_z$

$\hat{\Phi}$

$\mathbb{R}^k$
Heat Kernel Map

\[ (\mathcal{M}, g) \quad \text{dim} \mathcal{M} = d \]

\[ R_k \]

\[ R_z \]

\[ \hat{\Phi} \]

\[ \Phi \]

\[ \mathbb{R}^k \]

Saturday, May 23, 2009
Heat Kernel Map

\[ \Phi(x) = \left( R_z^d K_{R_z^2}(x, y_i) \right)_{i=1}^d \]

\[(\mathcal{M}, g) \quad \text{dim} \mathcal{M} = d\]
Heat Kernel Map

\[
\Phi(x) = (R_z^d K_{R_z^2}(x, y_i))_{i=1}^d
\]

\[(M, g) \quad \dim M = d\]
Heat Kernel Map

\[ \Phi(x) = (R_z^d K_{R_z^2} (x, y_i))_{i=1}^d \]
Theorem. Let \((\mathcal{M}, g), z \in \mathcal{M}\) and \((U, F)\) be as above, with the exception we now make no assumptions on the finiteness of the volume of \(\mathcal{M}\) and the existence of \(C_{\text{count}}\). Let \(\rho \leq r_U(z)\). Let \(p_1, \ldots, p_d\) be \(d\) linearly independent directions. There are constants \(c > 0\) and \(c', \kappa > 1\), depending on \(d\), \(c_{\text{min}}, c_{\text{max}}, \rho^\alpha \|g\|_\alpha\), \(\alpha\), and the smallest and largest eigenvalues of the Gramian matrix \(\langle p_i, p_j \rangle\) for \(i, j = 1, \ldots, d\), such that the following holds. Let \(y_i\) be so that \(y_i - z\) is in the direction \(p_i\), with \(c\rho \leq d_\mathcal{M}(y_i, z) \leq 2c\rho\) for each \(i = 1, \ldots, d\) and let \(t = \kappa^{-1}\rho^2\). The map

\[ x \mapsto (\rho^d K_t(x, y_1)), \ldots, \rho^d K_t(x, y_d)) \]

satisfies, for any \(x_1, x_2 \in B(z, \kappa^{-1}\rho)\),

\[ \frac{\kappa^{-1}}{c'\rho} d_\mathcal{M}(x_1, x_2) \leq \|\Phi(x_1) - \Phi(x_2)\| \leq \frac{\kappa c'}{\rho} d_\mathcal{M}(x_1, x_2). \]
Exit eigenfunctions

- 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.
- Little is understood about global properties of global eigenfunctions
- Behavior of eigenfunctions under perturbations of the graph
- Properties of eigenfunctions on graphs which are very different from sampled manifolds
- Relationships between eigenfunctions of different Laplacians
- 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.
- Little is understood about global properties of global eigenfunctions
- Behavior of eigenfunctions under perturbations of the graph
- Properties of eigenfunctions on graphs which are very different from sampled manifolds
- Relationships between eigenfunctions of different Laplacians
Enter multiscale analysis
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 random walks

We construct multiscale analyses associated with a diffusion-like process $T_t$ 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 $X \rightarrow X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_j \cdots$;

(ii) A coarsening (or compression) of the process $T_t$ at all time scales $t = t_j = 2^j$, $\{T_j = [T_{2^j}]_{\Phi_j}\}_j$, 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.
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:
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:

We fix \( \epsilon > 0 \), and we would like \( V_j \approx_{\epsilon} \langle \{ \phi_i : \lambda_i^{2^j+1} - 1 > \epsilon \} \rangle \).
Scheme for MRA construction

Notation: \([A]_{B_2}^{B_1} = B_2 AB_1^*\)
Let $T = D - \frac{1}{2} WD - \frac{1}{2} W^*$ as above be the $L_2$-normalized symmetric "random walk". The eigenvalues of $T$ and its powers "typically" look like this:

Notation: $[A]^{B_2}_{B_1} = B_2 AB_1^*$

Scheme for MRA construction

Dilations

Saturday, May 23, 2009
Scheme for MRA construction

Notation: \([A]_{B_1}^{B_2} = B_2 AB_1^*\)
Scheme for MRA construction

Notation: \[ [A]^{B_2}_{B_1} = B_2 A B_1^* \]
Compression step: more details

In order to compress the matrix $T$ we use “rank-revealing $QR$” decompositions. Fix $\epsilon > 0$.

$$T \Pi = QR = \begin{pmatrix} Q_{11} & Q_{12} \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix} \approx \epsilon \begin{pmatrix} Q_{11} & R_{11} \\ & R_{12} \end{pmatrix}$$

- $Q$ orthogonal, $R$ upper triangular, $\Pi$ permutation, $\|R_{22}\|_2 \approx \epsilon$
- $Q$ are the scaling functions $[\Phi_1]_{\Phi_0}$, $[R_{11} | R_{12}]$ is $[T]_{\Phi_0}^{\Phi_1}$, the compressed operator from fine to coarse scale.
- The number of columns $N_1$ of $Q_{11}$ (and of $R_{11}$) determines the dimension of the next coarse scale.
- The first $N_1$ columns of $\Pi$ select $N_1$ representative vertices on the graph.
Diffusion Wavelets on the circle

\[ T = \frac{1}{2} W D - \frac{1}{2} \]

The eigenvalues of \( T \) and its powers "typically" look like this:

Saturday, May 23, 2009
Diffusion Wavelets on the circle

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

Saturday, May 23, 2009
Diffusion Wavelets on the circle

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

Saturday, May 23, 2009
Example 1: on a manifold
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:
Consistence of multiscale r.w.'s

Let $G_j = \Phi_j$ be the graph whose vertices are the scaling functions at scale $j$. $T_j \approx [T^{2j+1}]_{\Phi_j}$ is a “random walk” (symmetrized) on $G_j$, a compressed version of $T^{2j}$ restricted to $V_j$. In fact:

$$T^{2j+1} f \approx P^*_V T_j P_V f$$

with $\approx$ meaning $\epsilon$-close in $L^2(G)$. We think of $(G_j, T_j)$ as a coarse version of $G$, constructed in such way that the random walk on $G_j$ is the random walk on $G$ at time $2^j$, compressed.
**Diffusion vs. Classical Wavelets**

<table>
<thead>
<tr>
<th></th>
<th>Classical MRA</th>
<th>Diffusion MRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dilation from scale $j$</td>
<td>By 2 (or dilation matrix $A$)</td>
<td>Apply $T^{2^j}$: both scaling and smoothing</td>
</tr>
<tr>
<td>to coarser scale $j + 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translations, downsampling</td>
<td>By $2^j k$, $k \in \mathbb{Z}^d$</td>
<td>Find well-conditioned set of scal. fcns.</td>
</tr>
<tr>
<td>Notion of “coarse”</td>
<td>$\tilde{V}_j$ is concentrated near 0</td>
<td>$V_j$ is approx. spanned by low-frequency eigenfunctions of $T$</td>
</tr>
<tr>
<td>Supports</td>
<td>Same (up to scale) for all scaling/wavelet fcns.</td>
<td>Different even at same scale</td>
</tr>
<tr>
<td>Vanishing moments</td>
<td>Orthogonality to poly,’s</td>
<td>Orthogonality to low-freq. eigenfcn,’s</td>
</tr>
<tr>
<td>Fast Wavelet Transform</td>
<td>$O(n)$</td>
<td>$O(n)$ in some cases, $O(n^2)$ in general</td>
</tr>
</tbody>
</table>
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.
Example 2: text documents
Example 2: text documents
Example 2: text documents
Example 2: text documents
### Example 2: text documents

<table>
<thead>
<tr>
<th>Scaling function</th>
<th>Document titles</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{2,3}$</td>
<td>Acid rain and agricultural pollution</td>
<td>nitrogen, plant, ecologist,</td>
</tr>
<tr>
<td></td>
<td>Nitrogen’s increasing impact in agriculture</td>
<td>carbon, global</td>
</tr>
<tr>
<td>$\phi_{3,3}$</td>
<td>Racing the Waves Seismologists catch quakes</td>
<td>earthquake, wave, fault,</td>
</tr>
<tr>
<td></td>
<td>Tsunami! At Lake Tahoe?</td>
<td>quake, tsunami</td>
</tr>
<tr>
<td></td>
<td>How a middling quake made a giant tsunami</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Waves of Death</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seabed slide blamed for deadly tsunami</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Earthquakes: the deadly side of geometry</td>
<td></td>
</tr>
<tr>
<td>$\phi_{3,5}$</td>
<td>Huntin prehistoric hurricanes</td>
<td>tornado, storm, wind,</td>
</tr>
<tr>
<td></td>
<td>Extreme weather: massive hurricanes</td>
<td>tornadoe, speed</td>
</tr>
<tr>
<td></td>
<td>Clearing the air about turbulence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>New map defines nation’s twister risk</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Southern twisters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oklahoma tornado sets wind record</td>
<td></td>
</tr>
</tbody>
</table>
Applications

- Multiscale approximation on data sets and graphs: applications to classification and regression problems on graphs.

- Multiscale/hierarchical representation of large graphs; visualization.

- Dictionary learning and efficient representation of data sets.

- Sparsity achieved through these multiscale representations can be exploited in various statistical settings - so far we obtain better than state-of-art results, both in terms of prediction rate and interpretability of results, for text document and gene array data [joint with J. Guinney, S. Mukherjee and P. Febbo].

Open problems & future dir.'s

- Construction of better diffusion wavelets
- Better understanding time-frequency analysis on graphs (e.g. Heisenberg principles, interactions between geometry and function spaces, etc...)
- Better understanding of multiscale approximation methods on graphs
- Deliver easy-to-use toolboxes for signal processing on graphs for use by non-experts
- Some of my current work: multiscale geometric tools for analysing the geometry of high-dimensional data sets, estimating intrinsic dimensionality, performing geometric wavelet analysis of the data; manifold-valued data.
Acknowledgments

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: NSF, ONR, Sloan Foundation, Duke.

THANK YOU!

www.math.duke.edu/~mauro