Diffusion-driven Multiscale Analysis on Manifolds and Graphs: top-down and bottom-up constructions

Arthur D Szlam, Mauro Maggioni, Ronald R Coifman, James C Bremer Jr.

Department of Mathematics, Program in Applied Mathematics, Yale University, 10 Hillhouse Ave, New Haven, CT, 06520

ABSTRACT

Classically, analysis on manifolds and graphs has been based on the study of the eigenfunctions of the Laplacian and its generalizations. These objects from differential geometry and analysis on manifolds have proven useful in applications to partial differential equations, and their discrete counterparts have been applied to optimization problems, learning, clustering, routing and many other algorithms.[1–7] The eigenfunctions of the Laplacian are in general global: their support often coincides with the whole manifold, and they are affected by global properties of the manifold (for example certain global topological invariants). Recently a framework for building natural multiresolution structures on manifolds and graphs was introduced, that greatly generalizes, among other things, the construction of wavelets and wavelet packets in Euclidean spaces.[8,9] This allows the study of the manifold and of functions on it at different scales, which are naturally induced by the geometry of the manifold. This construction proceeds bottom-up, from the finest scale to the coarsest scale, using powers of a diffusion operator as dilations and a numerical rank constraint to critically sample the multiresolution subspaces. In this paper we introduce a novel multiscale construction, based on a top-down recursive partitioning induced by the eigenfunctions of the Laplacian. This yields associated local cosine packets on manifolds, generalizing local cosines in Euclidean spaces.[10] We discuss some of the connections with the construction of diffusion wavelets. These constructions have direct applications to the approximation, denoising, compression and learning of functions on a manifold and are promising in view of applications to problems in manifold approximation, learning, dimensionality reduction.

Keywords: Multiscale analysis on manifolds, diffusion wavelets, biorthogonal wavelets, local cosines, diffusion on manifolds, Laplace-Beltrami operator.

1. INTRODUCTION

In this paper and its companion we discuss several recent and new ideas on how to perform multiscale analysis on manifolds and graphs.[11] While multiscale harmonic analysis has proven to be an extremely powerful tool across many fields and applications, it has been mostly confined to the Euclidean setting, in particular to local cosines and wavelets in $\mathbb{R}^n$, and to certain classes of groups. It is desirable to extend the multiscale tools of signal processing from the Euclidean setting to more general situations, in particular the case of manifolds and graphs. One of our main motivations is that in the context of data analysis, dimensionality reduction and learning, one is often interested in studying function whose domain can modeled as a manifold or a graph. Flexible tools for denoising, compressing, extract features of such functions seem of great importance.

The immediate question is how to efficiently organize a manifold or a graph, and functions defined on it, in a natural multiscale way. In the Euclidean setting there is a natural well-defined notion of time (or space) and frequency, and of localization in each of these domains. Multiscale analysis techniques based on wavelets or local cosines allow to (smoothly) split the time and frequency domains in multiscale fashions, construct atoms well-localized at arbitrary positions in time and frequency, and analyze functions with these atoms.

In the original diffusion wavelet paper a new paradigm for multiscale analysis on manifolds and graph is introduced, that greatly generalizes the Euclidean construction of wavelets.[8] There are three key ingredients in that construction. First of all, the idea to use powers of a diffusion operator on a manifold or graph to introduce

Send correspondence to Mauro Maggioni. E-mail: mauro.maggioni@yale.edu, Telephone: 1 203 432 8001

a notion of scales, both in time (or space) and in frequency, where the concept of frequency is induced naturally by the spectral theory of the operator. Hence, in this paradigm the focus shifts from dilation groups acting "geometrically" on the space, to diffusion semigroups acting on functions on the space. The second ingredient is the idea of enforcing a critical downsampling rate at each scale through a numerical constraint based on the effective rank of the powers of such operator. The third ingredient is an orthogonalization scheme that efficiently allows the computation of orthonormal, critically sampled bases at each scale. The method suggested in that paper is just one of many that are possible, and several new methods are being developed.

We review these ideas in detail in Section 3. This construction is "bottom-up": it starts with the atoms at the finest scale, and construct the atoms for analysis at coarser and coarser scales. It can be shown that the elements constructed have good localization properties in both time (or space) and frequency. In the companion paper we discuss a biorthogonal construction, which is shown to have better localization properties than the original orthogonal construction.[11]

Here we present a novel "top-bottom" construction, that yields on the one hand well-localized bases of "smoothed Haar wavelets", and on the other local cosines. The second eigenfunction of a diffusion on the manifold or graph is used to split the space into two pieces. Then each piece is recursively subdivided further, by using the second eigenfunction of the restriction of a diffusion operator to functions essentially supported on each piece. This yields a "dyadic" decomposition of the space, that in some sense generalizes the dyadic decomposition of Euclidean spaces. However very little is known of the properties of this decomposition, and in this paper we do not shed any light on the general properties it may have, that justify the results we obtain in practice. In the several examples we tried, it yielded extremely reasonable results.

In Section 2 we recall some basic material related to diffusion operators on manifolds and graphs, which is also relevant to the companion paper.

In Section 3 we briefly discuss the construction of diffusion wavelets, as a paradigm for bottom-up constructions.

In Section 4 we present several top-down constructions: first we discuss the "dyadic" partitioning of the space, then how such a decomposition can be used to construct Haar wavelets on a manifold or graph. We will explain how these Haar wavelets can then be smoothed and re-orthogonalized in such a way to generate localized smooth wavelets, and an associated multiresolution analysis. Finally we present a construction of local cosines, through general smooth projections induced by the "dyadic" decomposition constructed above.

2. DIFFUSION OPERATORS ON MANIFOLDS AND GRAPHS

In this section we introduce some of the definitions and results related to diffusion on manifolds and graphs which are the foundation of the constructions we present in later sections. These topics are classical: in the case of manifolds, the Laplacian, the heat kernel, and the corresponding eigenfunctions have been studied for decades. While a lot is understood, many open problems still exist in the geometry of the eigenfunctions of the Laplacian; for instance regarding the structure of the nodal lines and of the nodal sets. In the case of graphs, spectral graph theory[7] has had successful applications, for example to approximation algorithms in computer science. Recently diffusions on manifold, graphs, and data sets have been shown to be useful for tasks in clustering, data analysis, learning and several others.[1–6]

2.1. The Laplacian on a Riemannian manifold

Let $(\mathcal{M}, g)$ be a smooth Riemannian manifold. The Laplacian operator is natural operator on graphs and manifolds. It is related, via the heat equation, to natural random walks (Brownian motion) on the graph or manifold. In local coordinates, the expression of the Laplace-Beltrami acting on a smooth function f with compact support is

$$\Delta f = -\frac{1}{\sqrt{\det g}} \partial_j (g^{ij} \sqrt{\det g}\, \partial_i f)\,. \tag{1}$$

Intrinsically, it can be defined by introducing the operators div and grad on the Riemannian manifolds, and letting $\Delta f = -\mathrm{div}(\mathrm{grad} f)$. The Laplacian is essentially self-adjoint, and can be extended to an unbounded

operator Δ on $\mathcal{L}^2(\mathcal{M})$, where the measure on $\mathcal{M}$ is the natural Riemannian volume element. The heat equation on $(\mathcal{M}, g)$, is

$$\frac{\partial}{\partial t}u(t,x) = \Delta u\,, \tag{2}$$

together with some initial condition specifying u at time $t = 0$, and some boundary conditions, if M has a boundary. Dirichlet boundary conditions impose $u|_{\partial M} = 0$, Neumann boundary conditions impose $\frac{\partial}{\partial \nu}u|_{\partial M} = 0$, where ν is outward normal, and mixed boundary conditions are a linear combination of Dirichlet and Neumann conditions. The heat kernel is defined as the $\mathcal{C}^\infty$ function $H_t(x,y) : \mathbb{R}_+ \times M \times M \to \mathbb{R}$ such that

$$\begin{cases} (\partial_t + \Delta_x)H_t(x,y) = 0 \\ \lim_{t\to 0^+} \int_M H_t(x,y)f(y)dy = f(x)\,. \end{cases} \tag{3}$$

For a manifold with boundary, we will require $H_t(x,y)$ to satisfy the boundary conditions (Dirichlet, Neumann, or mixed) for each t. One can show that there exists exactly one such function (even when imposing less stringent smoothness assumptions), which is also the kernel of the operator $e^{-t\Delta}$:

$$e^{-t\Delta}f(x) = \int_M H_t(x,y)f(y)dy\,.$$

The eigenfunctions of the Laplacian will play a very important role in the sequel. Hodge's Theorem says that, at least when $\mathcal{M}$ is compact, there exists a complete orthonormal basis of eigenvectors $\{\xi_i\}_{i\geq 0}$ of Δ with corresponding eigenvalues $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_j \leq \ldots$ which accumulate only at infinity. One can expand the heat kernel on these eigenvectors:

$$H(t,x,y) = \sum_{i\geq 0} e^{-\lambda_i t}\xi_i(x)\xi_i(y)\,, \tag{4}$$

with convergence at least $\mathcal{L}^2(\mathcal{M})$, but in fact, when $\mathcal{M}$ is compact, even uniform, with uniform convergence of the derivatives as well. One way to prove this result is by proving first the compactness of the heat kernel operator, and then studying its properties as it acts on smooth functions.

The are a number of related references for the interested reader that cover classical material regarding the Laplacian and the heat kernel on a Riemannian manifold.[12–19]

2.2. The Laplacian on a graph

Let G be a weighted, non-oriented, graph, with vertex set $V(G)$, edge set $E(G)$ and weights (on the edge set) $W(G)$. The weights are typically assumed to be nonnegative and symmetric, in the sense that $w(x,y) = w(y,x)$ for every edge $\overline{xy}$. The degree d_x of a vertex x is defined to be:

$$d_x = \sum_{x\sim y} w(x,y)\ .$$

The volume of the graph is defined as $\operatorname{vol} G = \sum_x d_x$. The *Laplacian* on $G = G(V,E,W)$ is the operator L acting on compactly supported functions on G by

$$Lf(x) = \sum_{y\sim x} (f(x) - f(y))\, w(x,y)\ . \tag{5}$$

If $w(x,y)$ is "local", in the sense that, for each x, there are only few values of y such that $w(x,y)$ is nonzero, then the summation in (5) is "local". The Laplacian can be represented on the basis $\{\delta_x\}_{x\in G}$ of Kronecker δ-functions by the matrix

$$L(x,y) = \begin{cases} d_v - w(x,y) & \text{if } x = y, \\ -w(x,y) & \text{if } x \text{ and } y \text{ areadjacent}, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

One usually introduces the matrices D and W defined as follows. D is the diagonal matrix whose entry $D_{xx} = d_x$, and W is the matrix of weights, $W_{xy} = w(x, y)$. Then clearly

$$L = D - W \ . \tag{7}$$

It is easy to see that L is positive semi-definite. Very often, one considers the *normalized Laplacian* $\mathcal{L}$, which is defined by

$$\mathcal{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}} \ . \tag{8}$$

We are of course assuming here that $d_x > 0$ for every $x \in G$ (no isolated points). It acts on functions on G by

$$\mathcal{L}f(x) = \frac{1}{\sqrt{d_x}} \sum_{x \sim y} \left(\frac{f(x)}{\sqrt{d_x}} - \frac{f(y)}{\sqrt{d_y}} \right) \ . \tag{9}$$

The Spectral Theorem applied to $\mathcal{L}$ guarantees the existence of an orthonormal basis of eigenvectors $\{\xi_i\}$ corresponding to eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_i \leq \ldots$, so that

$$\mathcal{L}\xi_i = \lambda_i \xi_i \ .$$

The spectral decomposition of L is then given by $L(D^{-\frac{1}{2}}\xi_i) = \lambda_i (D^{-\frac{1}{2}}\xi_i)$, so that $T^{\frac{1}{2}}\xi_i$ are the eigenvectors of L, and $\sigma(L) = \sigma(\mathcal{L})$.

2.3. Eigenvectors of the Laplacian and Fourier analysis

The eigenvectors ξ_i of the Laplacian on a space X, which is either a Riemannian manifold or a graph, form a natural orthonormal basis of functions for analysis on the space. In fact they are the natural generalization of Fourier modes on Euclidean spaces, which are (generalized) eigenfunctions of the Euclidean Laplacian.

The eigenfunction ξ_i has a frequency related to the corresponding eigenvalue λ_i. For example this can be measured by looking at the L^2 norm of the gradient, which is equal to λ_i since $\int_{\mathcal{M}} ||\nabla \xi_i||^2 = \int_M \langle \Delta \xi_i, \xi_i \rangle$. Also, eigenfunctions corresponding to higher eigenvalues have more nodal domains (but there are still open questions regarding estimates on the number of nodal domains for high frequency eigenfunctions). A *nodal domain* is a maximal set on which the eigenfunction does not change sign. They are separated by *nodal lines*. For example it is known that the second eigenfunction, if the space is connected, has exactly two nodal domains.

Smoothness spaces, such as Sobolev spaces, can be characterized by the rate of decay of the coefficients of a function in the class onto the eigenfunctions ξ_i, appropriately weighted by (powers of) the eigenvalues λ_i.

Fourier analysis on manifolds and graphs performed using the eigenfunctions of the Laplacian has some of the same drawbacks of Fourier analysis on Euclidean spaces. Because the Fourier modes are in general global, they can used to detect global smoothness, and are going to be highly inefficient for representing functions with different smoothness properties in different regions. They cannot be expected to perform very well in applications such as compression and denoising of functions.

Wavelets or local cosines, or some other type of multiscale analysis would be much better suited for these applications. In the rest of the paper we will present a generalization of these constructions to manifolds and graphs.

3. BOTTOM-UP CONSTRUCTIONS: DIFFUSION WAVELETS

A framework for multiscale analysis on graphs and manifold is introduced, and orthonormal diffusion wavelets, for performing such analysis efficiently, have been recently constructed.[8]

There are three key ingredients in that construction. First of all the idea to use the diffusion on the manifold or graph to introduce a notion of scales. Secondly a way of enforcing the critical rate of downsampling at each scale through a numerical constraint. Thirdly, an orthogonalization scheme that rather efficiently allows the computation of orthonormal, critically sampled bases at each scale.

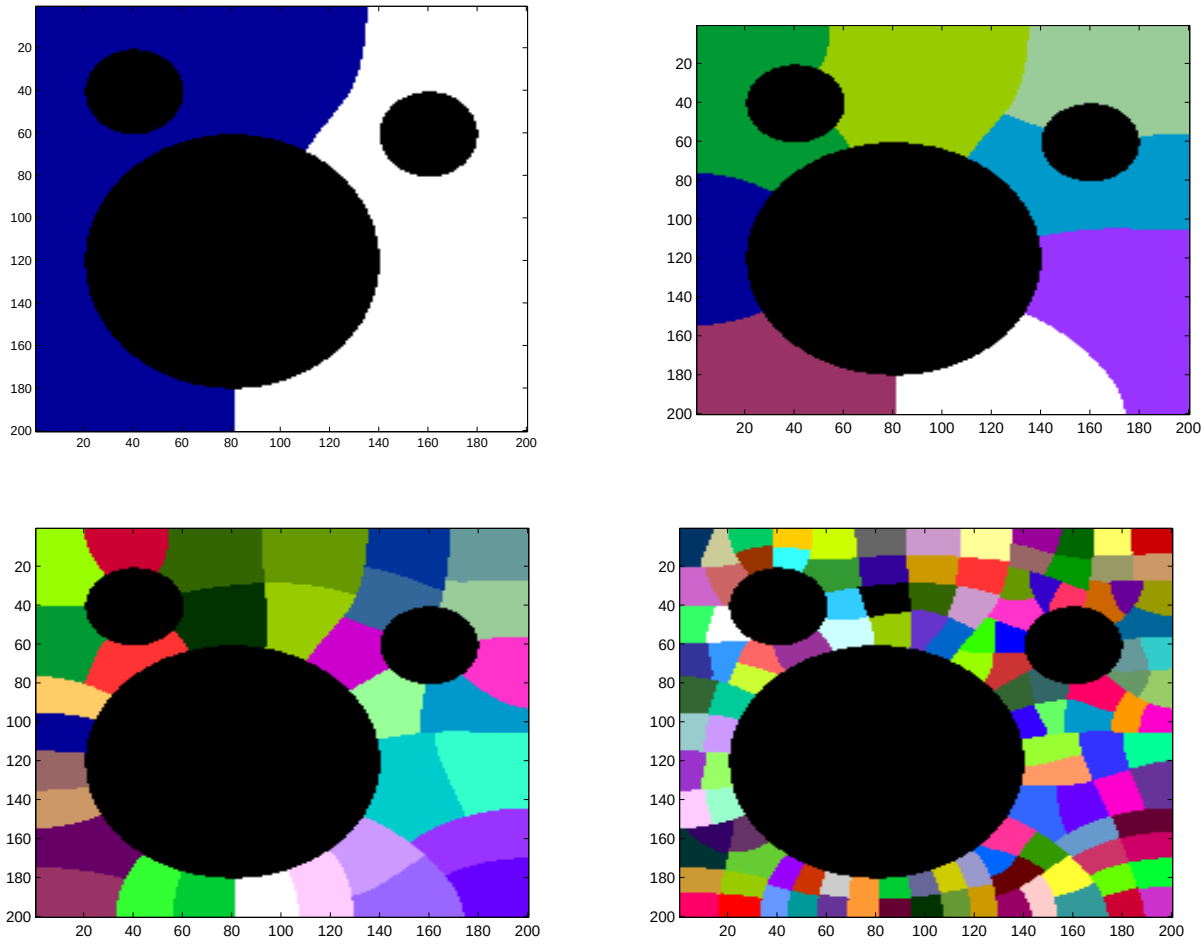

Figure 1. levels 2, 4, 6, and 8 in the decomposition of a planar domain with holes, uniformly sampled at 28,000 points.

For a summary of the original construction we refer the reader to the paper,[8] and for the new biorthogonal construction we refer the reader to the companion paper.[11] Here we make the following remarks. First of all the construction is bottom-up: the input to the construction is a set of orthonormal atoms Φ_0 at the finest scale of interest, together with a diffusion semigroup (e.g. the heat semigroup associated with the heat kernel on the manifold or graph). Dyadic powers of the generator T of the diffusion semigroup are used as dilations, acting on the given family of fine-scale atoms. The construction proceeds as follows: T is applied to Φ_0, and an orthonormal basis Φ_1 is constructed from $T\Phi_0$, that spans, up to a pre-specified precision, the same subspace. Observe that the cardinality of Φ_1 may be smaller than the cardinality of Φ_0, since, up to the pre-specified precision, T may not be full rank. Then T is restricted to the span of Φ_1, squared, and the family $T^2\Phi_1$ is considered. Observe that the coordinates are now with respect to Φ_1: both T^2 and $T^2\Phi_1$ are represented on the basis Φ_1. Once again an orthonormal basis Φ_2 for the span (up to the pre-specified precision) of $T^2\Phi_1$ is constructed. The cardinality of Φ_2 may be smaller than the cardinality of Φ_1, since T^2, up to the pre-specified precision, may not be full rank. We proceed in this fashion until the coarsest scale is reached. Observe that the rank of powers of T, up to any pre-specified precision, is small, because T is a smoothing operator, and hence the numerical rank of its powers is small.

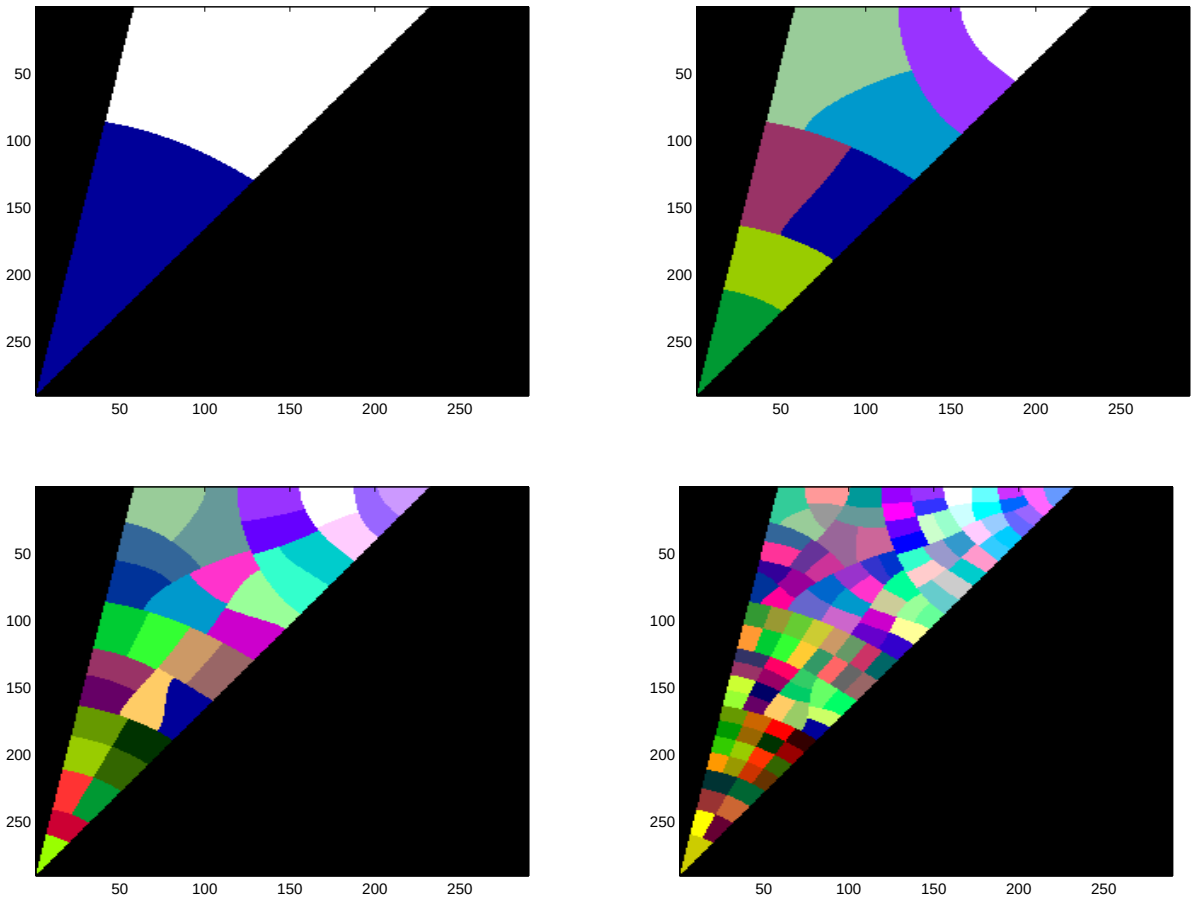

Figure 2. levels 2, 4, 6, and 8 in the decomposition of a wedge, uniformly sampled at 25,000 points.

4. TOP-DOWN CONSTRUCTION

4.1. Decomposition

The first ingredient of the top down construction is a "dyadic" recursive decomposition of the manifold, obtained by using the nodal lines of Neumann eigenfunctions computed recursively on the pieces of the decomposition at the previous scale. We will find the (any) first nonconstant eigenfunction of the Neumann Laplacian on the manifold, and decompose it into the nodal domains, i.e. the set of points where the eigenfunction is positive, and the set of points where the eigenfunction is negative. The Courant Nodal Domain theorem guarantees that there are two nodal domains. The first non-constant Neumann eigenfunction on each nodal domain can then be considered and the process iterated to form our decomposition. At the finest scale of the subdivision we actually cut using principal components rather than eigenfunctions of the Laplacian.

Graph theorists have considered this kind of decomposition in the past. For example the eigenfunction has served as an approximation of the normalized min-cut of a graph. In particular this decomposition is natural if the goal is to obtain pieces that have small boundary-area to volume ratio: repeatedly splitting along the normalized min-cut is the greedy algorithm for creating a decomposition with sets whose area/volume ratio should not be too bad, and the division according to the eigenfunction of the Laplacian is a readily computable guess at the min-cut.[20] However, it can happen that the cut chosen by the Laplacian will be significantly worse than ideal with respect to the normalized cut criterion. Moreover, we are not able to guarantee that this greedy algorithm yields a subdivision whose pieces do have small boundary-area to volume ratios, nor are we aware of

any results that may guarantee this. In fact rather "artificial" counterexamples show that this may not be the case, but it seem hard to characterize precisely what "artificial" means in this context.

Though the eigenfunction is only a guess at the min-cut, there is an exact description of the decomposition which is intuitively very appealing: 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 $\Gamma = \partial D_i - \partial D$. To see this, first we notice that Γ is the common boundary between the two sub-domains that have minimum (equal) mixed Neumann-Dirichlet eigenvalue, which follows from comparing the variational characterization of the mixed problem to the variational characterization of the Neumann problem. Now for any sub-domain of D, if $H_t(x,y) = \sum_{n=0}^{\infty} e^{-\lambda_n t}\phi_n(x)\phi_n(y)$ is the mixed kernel, the probability of no escape at time t is given by

$$P(B_x(t) \in D_x) = \int_{D_x} H_t(x,y)dy,$$

so

$$\lim_{t\to\infty} e^{-\lambda t} P(B_x \notin D_x) = C\phi_0(x),$$

where $C = \int_{D_x} \phi_0(y)dy$ (this is proved, but not stated as such in[21]). Thus the division according to the eigenfunction is like maximizing the volume to boundary area ratio as seen by a Brownian particle.

It seems important that we take Neumann eigenfunctions. The main reason for this is that this decomposition is the greedy algorithm for subdividing a manifold so the subsets have large Neumann second eigenvalue, and we expect the second eigenvalue to be a decent indicator of the boundary regularity of the domain (as opposed, for example, to the Dirichlet first eigenvalue, which is essentially unaffected by adding thin sets to the boundary). Secondly, given a sampled manifold with boundary, the Neumann condition is easy to model. There are also some advantages that help with the construction of localized bases on the subsets. For example if the nodal line of the first Neumann eigenfunction meets the boundary of the domain away from a corner, it will do so at a right angle.

Experimentally this works extremely well on sampled surfaces and domains in 2 and 3 dimensions. We make two wild conjectures for domains in $\mathbb{R}^2$: There is a constant C_D depending on the initial domain so that $\lambda_{D_s}|D_s| > C_D$ for all subdivisions s, and if we specify that we take the eigenfunction with shortest nodal line if there is more than one (for domains in $\mathbb{R}^2$ there are at most two[21]), then the sets limit to polar rectangles, squares, or right triangles. Note that this decomposition is not dyadic in the sense that the areas of the subsets at a given level are not comparable; for example, take a narrow polar wedge. The cut will always take off a bit of the sharp corner, and that piece will always have less than half the area of its parent. In Figure 1 we show the subdivision, at different scales, for square domain with round holes in $\mathbb{R}^2$ uniformly sampled at $28,000$ points. In Figure 2 we show the subdivision, at different scales, for a planar wedge in $\mathbb{R}^2$ uniformly sampled at $25,000$ points. Note how the sets near the sharp corners become polar sectors.

4.2. Classical local cosine and best basis

The classical construction of local cosine bases,[10] say on the interval $[-1,1]$, allows to localize Fourier analysis to windows, yielding an orthonormal basis $\{e_{\tau,\nu}\}_{\tau\in T,\nu\in F}$, where T is a set of locations and F a set of frequencies, and $e_{\tau,\nu}$ is well-localized in the time-frequency plane around (τ,ν). Roughly speaking $e_{\tau,\nu}$ is obtained by carefully windowing around τ a pure Fourier mode of frequency ν, but in such a way to preserve orthogonality between different elements. This is achieved through the careful construction of two projections on $L^2([-1,1])$ which smoothly split L^2 into functions essentially supported to the right interval, and functions essentially supported to the left interval. The data necessary for the construction of the projections is a rising cutoff function s that satisfies $s^2(t) + s^2(-t) = 1$, $s = 0$ for $t < -\epsilon$, and $s = 1$ for $t > \epsilon$. The projections are given by

$$P_1 f(t) = s^2(t)f(t) + s(t)s(-t)f(-t)$$

$$P_2 f(t) = s^2(-t)f(t) - s(t)s(-t)f(-t).$$

This procedure can be iterated, yielding a multiscale family of subspaces of L^2 so that any two subspaces at a given scale are orthogonal, the projections preserve smoothness, and each function in a given subspace is compactly supported, mostly on a specific interval of the proper scale.

Fast "best basis" algorithms based on dynamic programming exist that allow to choose, for a given signal or family of signals, the set of local cosine subspaces for better compressing or denoising or performing a discrimination task.

4.3. Smoothed Haar Bases

Once a (dyadic) decomposition of the manifold is given, we can attempt to build bases with elements localized on a few subsets of that decomposition. The simplest basis we can pick consists of Haar scaling functions and wavelets. The scaling spaces for the Haar wavelets are spanned by the functions constant on the subsets of the decomposition at a given scale.

This makes it possible to downsample a manifold in the following way. Suppose we are given a set of coordinate functions for the manifold. These functions can be expanded on the Haar wavelets. Removing the small scale coefficients generates a coarser version of the coordinate functions. This thresholding corresponds to subsampling the manifold. The effectiveness of this operation is, of course, dependent on the quality of the partition, about which very little is known. As mentioned before, although we do not expect the volumes of the partition to be uniform, we expect the boundaries of the cubes to be relatively regular. If necessary, one could correct for the difference in volumes with a stopping time. The algorithms for the Haar functions are fast, i.e. $n \log n$ (the constant depends on the relative volumes of the subsets at a given level in the partition).

The major problem with the Haar functions is of course that they are not smooth. We can smooth the functions using a heat operator; in order to obtain a uniform smoothness for the modified Haar system, it is necessary to apply different powers of T at different scales. More precisely, we apply $T^{2^{j-1}}$ to smooth the Haar functions at scale j. Of course these smoothed Haar functions are no longer orthogonal.

It is also quite feasible to do the following non-linear analysis of a function f: first, find the Haar transform of f, then the m largest coefficients, then orthonormalize the corresponding smoothed Haar wavelets. The entire procedure takes $\sim mn \log n$ operations, essentially the same as finding m eigenfunction coefficients. In the example below, we orthonormalize the m smoothed Haar wavelets by pivoting by the size of the coefficient of f onto the original Haar functions. For functions with energy localized spacially, this can lead to better compression than using the basis of eigenfunctions, as we show in example in Figure 3. In Figure 4 we represent 3 smoothed Haar functions intervening in the reconstruction of the function in the example of Figure 3.

The order in which the smoothed Haar functions are orthonormalized can in general be important, in the sense that it may affect the decay, the size of the support, and the smoothness properties of the resulting orthonormal basis. If we orthogonalized from the bottom up, we would have much better control of the supports of the functions, but the functions at large scale would tend to be less smooth.

4.4. Local cosines on manifolds and graphs

We now build local cosines with respect to the eigenfunction decomposition. The one-dimensional local cosine construction utilizes two projections on $L^2([-1,1])$ based on a smooth rising cutoff function s that satisfies $s^2(t) + s^2(-t) = 1$, $s = 0$ for $t < -\epsilon$, and $s = 1$ for $t > \epsilon$ (call$-\epsilon < t < \epsilon$ the action region). The projections are given by

$$\begin{aligned} P_+ f(t) &= s^2(t)f(t) + s(t)s(-t)f(-t) \\ P_- f(t) &= s^2(-t)f(t) - s(t)s(-t)f(-t)\,. \end{aligned} \tag{10}$$

These projections take smooth functions to smooth functions. This process can be iterated on subintervals, although becomes problematic when regions start to intersect.

Given a manifold with boundary M and submanifolds $M_+ = \phi > 0$, and $M_- = \phi < 0$ where ϕ is the first eigenfunction of the Neumann Laplacian, if $|\nabla\phi| \neq 0$ on the nodal set, we can directly generalize the first

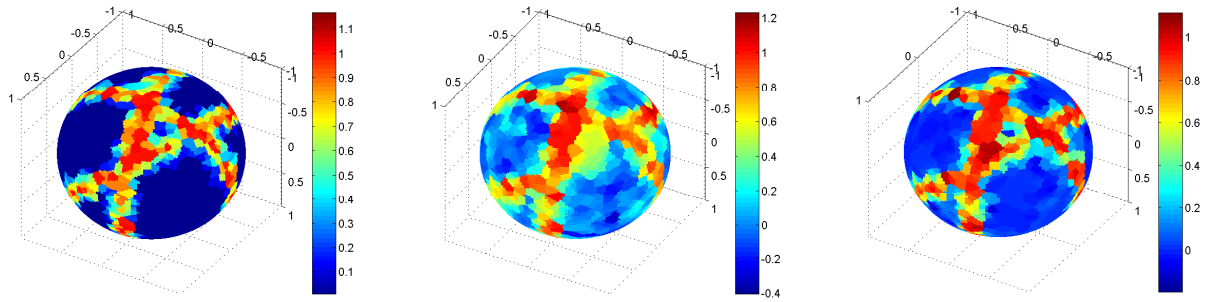

Figure 3. A function on a sphere sampled randomly at 2000 points: a sum of 3 random great circles smoothed by the heat operator. In the next window, reconstructed by the top 256 heat eigenfunctions. Finally, reconstructed by 256 adapted smooth Haar functions. The L^2 reconstruction errors are 27% and 16% respectively.

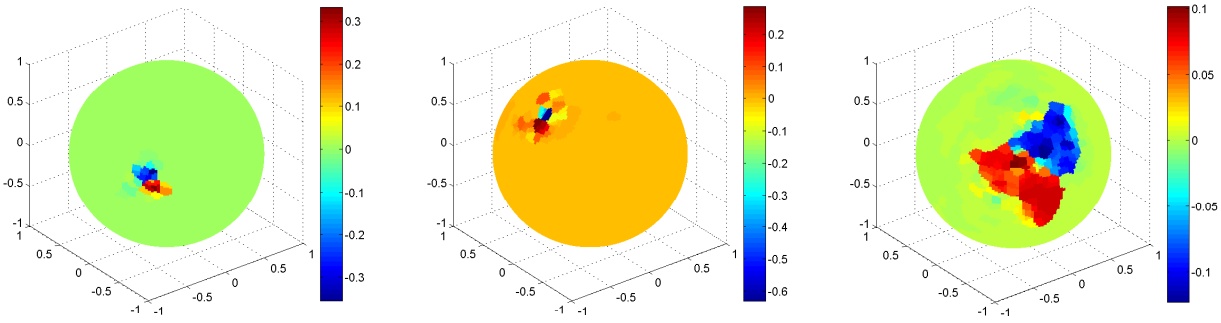

Figure 4. Three smoothed Haar functions on a sphere sampled randomly at 2000 points.

iteration of the one-dimensional construction. The first step is to construct a reflection that replaces the natural map $t \mapsto -t$ on the interval. Our reflection r, defined in a neighborhood of the nodal set, takes a point x to a point y such that $\phi(x) = -\phi(y)$, and x and y lie on the same trajectory orthogonal to the nodal line. We replace in (10) the natural one-dimensional reflection given by $t \mapsto -t$ with r (note that choosing a Neumann eigenfunction means that the orthogonal trajectories are parallel to the boundary when $|\nabla\phi| \neq 0$ on the nodal set), and build cutoff functions using ϕ as a coordinate, normalized so the reflection r is unitary. Then everything from the one-dimensional case carries over exactly.

However, to continue subdividing as in one dimension is difficult, because many subsets can share common boundaries, and the sequential reflections do not necessarily interact well. One solution to this problem has been given by Kovacevic and Bernardini.[22,23] They suppose a group of transformations acts on all the common boundaries and use representations of the group to generalize the construction of the projections. It turns out to be difficult to implement this solution because it is hard to find with a group of transformations unitarily relating the various neighboring sets. In fact, in the discrete case, even finding a single unitary reflection can be quite difficult — assuming the boundary region curves, one does not even always exist, because locally there can be more points on one side of the nodal line than the other. To overcome these obstacles, suppose we have a good reflection, and have constructed the projections. Consider the operator $C = P_+ - P_-$. Let K be a smoothing operator, for example the heat operator. Note that $C^2 = I$, hence C commutes with $KC + CK$, making $KC + CK$ and C simultaneously diagonalizable. The joint eigenfunctions of these operators are smooth bases primarily supported on one the two nodal sets, and decaying in an "action region" around the nodal line. Our approach, then, will be to find a smoothing operator whose positive eigenvalues correspond to eigenfunctions supported on $\phi < \epsilon$, and whose negative eigenvalues correspond to eigenfunctions supported on $\phi > -\epsilon$.

In our actual implementation, we do not even try to find a reflection. We proceed instead in the following manner:

- pick functions $\{v_i\}$, each supported on one of the nodal regions, but not necessarily smooth;
- form a cutoff operator for each side;
- smooth the functions $\{v_i\}$, for example by applying a diffusion operator such as the heat operator;
- multiply by the cutoff operator to obtain functions $\tilde{v}_i$;
- proceed as if $\{\tilde{v}_i\}$ were an orthonormal system, and construct the symmetric matrix $\Sigma\epsilon(i)\tilde{v}_i(j)\tilde{v}_i(k)$, where $\epsilon(i)$ is 1 or -1 depending on which nodal regions contains the support of $\tilde{v}_i$;
- compute the orthonormal basis of eigenvectors of Σ, and corresponding orthogonal projections P_+ and P_- onto the positive and negative eigenspaces;
- find a smooth set of eigenfunctions corresponding to these nodal sets by diagonalizing $K(P_- - P_+) + (P_- - P_+)K$, where again K is a (compactly supported) smoothing operator.

Finally, we illustrate a method for iterating this algorithms in order to obtain a multiscale decomposition. Suppose we are given the first projections P_+ and P_-, and an action region A around the nodal line. Let M_+ be the support of the image, under P_+, of the functions . Consider the nodal sets $\widetilde{M_{++}}$ and $\widetilde{M_{+-}}$, and action region A_+. Form $\widetilde{P_{++}}$ and $\widetilde{P_{+-}}$ as above. Note that these operators are defined on $L^2(M_+)$, and do not necessarily leave the subspace $P_+(L^2(M))$ invariant. So we form the operator $T = P_+\widetilde{P_{++}}K\widetilde{P_{++}}P_+$, where K is a smoothing operator with compactly supported kernel. The eigenfunctions of T are supported on $\widetilde{M_{++}} \bigcup A_+ \bigcup A$ spread by K, and so we obtain the next projection P_{++}, by projecting onto the eigenfunctions with positive eigenvalues. P_{+-} is just $P_+ - P_{++}P_+$, and M_{++} and M_{+-} are the union of the respective nodal sets and the action region A_+. The same procedure is carried out on M_-. The total cost for a subspace is $\sim n^3$, making this a somewhat expensive algorithm. However, the constant can be made very small by only working in the action regions, and using relatively narrow action regions (which of course makes the basis less smooth). It is also unsatisfactory in the sense that although for any subset in the decomposition there is localization away from the 'cores' of other subsets, as one iterates, the local cosines can be supported on all the action regions of their ancestors, because in solving the eigenproblem, we loose control of the support where the projections are not both the identity. However, it may be possible to recursively apply the construction of the smooth projections to the action region, and then the action region of the action region, etc, to keep descendants from living on ancestor's action regions, and to speed up the algorithm.

5. EXAMPLE

We consider a sphere sampled randomly uniformly at 2000 points, and construct local cosines with the algorithm described above. We plot in Figure 5 the resulting local cosine at level 4. They are localized oscillatory functions as desired. We compare their smoothness with that of the 6th heat eigenfunction, which is not localized at all.

6. ACKNOWLEDGEMENTS

MM is partially supported by NSF Grant DMS 0512050.

REFERENCES

1. R. Coifman and S. Lafon, "Geometric harmonics," tech. rep., Yale University, Dept. Comp. Sci., 2003.
2. S. Lafon, *Diffusion maps and geometric harmonics*. PhD thesis, Yale University, Dept of Mathematics & Applied Mathematics, 2004.
3. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comp. Harm. Anal.* , 2004.
4. R. Coifman and S. Lafon, "Geometric harmonics," *Appl. Comp. Harm. Anal.* , 2004 2004. Submitted.
5. M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation* **6**, pp. 1373–1396, June 2003.

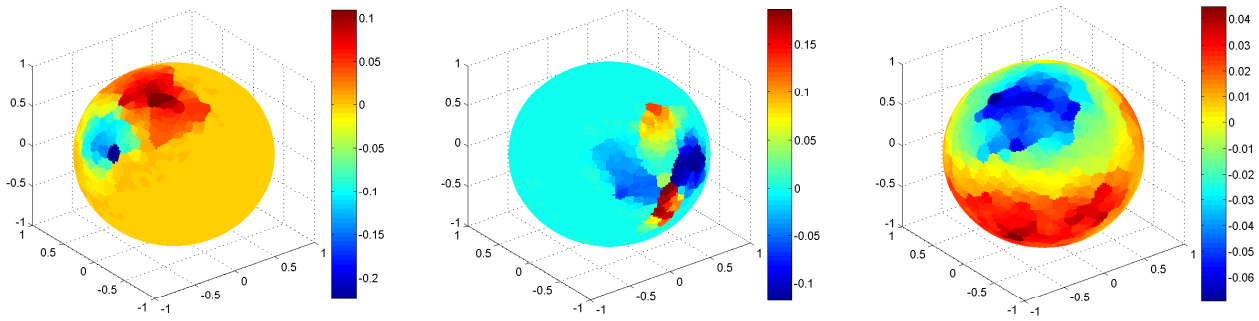

Figure 5. 2 local cosine functions at level 4, and the 6th heat eigenfunction to compare smoothness, all on a sphere randomly sampled at 2000 points, according to the uniform distribution.

6. P. Niyogi and M. Belkin, "Semi-supervised learning on Riemannian manifolds," Tech. Rep. TR-2001-30, University of Chicago, Computer Science Dept, Nov. 2001.
7. F. Chung, *Spectral Graph Theory*, no. 92, CBMS-AMS, May 1997.
8. R. Coifman and M. Maggioni, "Diffusion wavelets," *Tech. Rep. YALE/DCS/TR-1303, Yale Univ., Appl. Comp. Harm. Anal.* , Sep. 2004. to appear.
9. J. Bremer, R. Coifman, M. Maggioni, and A. Szlam, "Diffusion wavelet packets," *Tech. Rep. YALE/DCS/TR-1304, Yale Univ., Appl. Comp. Harm. Anal., submitted* , Sep. 2004.
10. R. Coifman and Y. Meyer, "Remarques sur l'analyse de Fourier à fenetre," *C.R. Acad. Sci. Paris* , 1991.
11. M. Maggioni, A. D. Szlam, R. R. Coifman, and J. C. B. Jr, "Biorthogonal diffusion wavelets for multiscale representations on manifolds and graphs," August 2005. SPIE Wavelet XI.
12. S. Rosenberg, *The laplacian on a Riemannian manifold*, vol. 31 of *Student Texts*, Cambridge University Press, London Mathematical Society.
13. J. M. Lee, *Riemannian manifolds: An introduction to curvature*, Springer, 1997.
14. P. Li and S. T. Yau, "On the parabolic heat kernel of the Schrödinger operator," *Acta Math.* **156**, pp. 153–201, 1986.
15. E. P. Hsu, "Estimates of derivatives of the heat kernel on a compact Riemannian manifold," *Proc. Amer. Math. Soc.* **127**, pp. 3739–3744, May 1999.
16. S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*, Springer-Verlag, Berlin, 1987.
17. S. Y. Cheng and P. Li, "Heat kernel estimates and lower bounds of eigenvalues," *Comment. Math. Helv.* **56**, pp. 327–338, 1981.
18. N. Sidorova, O. G. Smolyanov, H. v Weizsäcker, and O. Wittich, "Brownian motion close to submianifold of Riemannian manifolds." preprint, 2003.
19. J. Wang, "Globalheatkernelestimates," *Pacific Jour. of Math.* **178**(n), pp. 377–398, 1997.
20. J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Tran PAMI* **22**(8), pp. 888–905, 2000.
21. R. Bañuelos and K. Burdzy, "On the "hot spots" conjecture of J. Rauch," *J. Funct. Anal.* **164**(1), pp. 1–33, 1999.
22. R. Bernardini and J. Kovačević, "Designing local orthogonal bases on finite groups. I. Abelian case," *J. Fourier Anal. Appl.* **6**(1), pp. 1–23, 2000.
23. R. Bernardini and J. Kovačević, "Designing local orthogonal bases on finite groups. II. Nonabelian case," *J. Fourier Anal. Appl.* **6**(2), pp. 207–231, 2000.