Biorthogonal Diffusion Wavelets for Multiscale Representations on Manifolds and Graphs

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

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

ABSTRACT

Recent work by some of the authors presented a novel construction of a multiresolution analysis on manifolds and graphs, acted upon by a given symmetric Markov semigroup $\{T^t\}_{t\geq 0}$, for which T^t has low rank for large t.[1] This includes important classes of diffusion-like operators, in any dimension, on manifolds, graphs, and in non-homogeneous media. The dyadic powers of an operator are used to induce a multiresolution analysis, analogous to classical Littlewood-Paley[14] and wavelet theory, while associated wavelet packets can also be constructed.[2] This extends multiscale function and operator analysis and signal processing to a large class of spaces, such as manifolds and graphs, with efficient algorithms. Powers and functions of T (notably its Green's function) are efficiently computed, represented and compressed. This construction is related and generalizes certain Fast Multipole Methods,[3] the wavelet representation of Calderón-Zygmund and pseudo-differential operators,[4] and also relates to algebraic multigrid techniques.[5] The original diffusion wavelet construction yields orthonormal bases for multiresolution spaces $\{V_j\}$. The orthogonality requirement has some advantages from the numerical perspective, but several drawbacks in terms of the space and frequency localization of the basis functions. Here we show how to relax this requirement in order to construct biorthogonal bases of diffusion scaling functions and wavelets. This yields more compact representations of the powers of the operator, better localized basis functions. This new construction also applies to non self-adjoint semigroups, arising in many applications.

Keywords: Biorthogonal wavelets, Laplace-Beltrami operator, graph Laplacian, multiscale analysis on manifolds.

1. INTRODUCTION

In the companion paper we discuss novel approaches for multiscale analysis on manifolds and graphs.[7] There we divide the approaches into two categories, bottom-up and top-bottom, and discuss a novel top-bottom construction leading to local cosines on a manifold or graph.

In this paper we discuss in detail a novel bottom-up construction that generalize orthogonal diffusion wavelets,[1] leading to biorthogonal diffusion wavelets. In order to perform multiscale analysis on a manifold or graph, one would like to define dilations, that induce scales and scaling spaces, and to be able to interpret them both spatially and in terms of "frequency". One would also like to define some sub-sampling rule in order to "critically sample" each of the scaling spaces: the number of atoms at each scale should be smaller the coarser the scale, but in some sense still be "enough" to represent faithfully any function "at that scale". Finally, one would also like to be able to efficiently compute and represent the constructed multiscale structure, in terms of filters that allow to encode each scale in terms of the previous scale, and to efficiently analyse and reconstruct functions in terms of the constructed scaling function and wavelets.

A novel framework for multiscale analysis on manifolds and graph was recently introduced,[1] that greatly generalizes wavelet analysis to manifolds and graphs, and even more general spaces, and coherently addresses the issues above. It is related to a circle of ideas centered around the use of diffusion operators on manifolds, graphs, and general data sets.[8–13]

As far as dilations are regarded, the main idea is to shift the attention from the geometry of the space to the geometry of certain families of functions on the space.[1,14] Instead of considering groups (or semigroups) acting

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

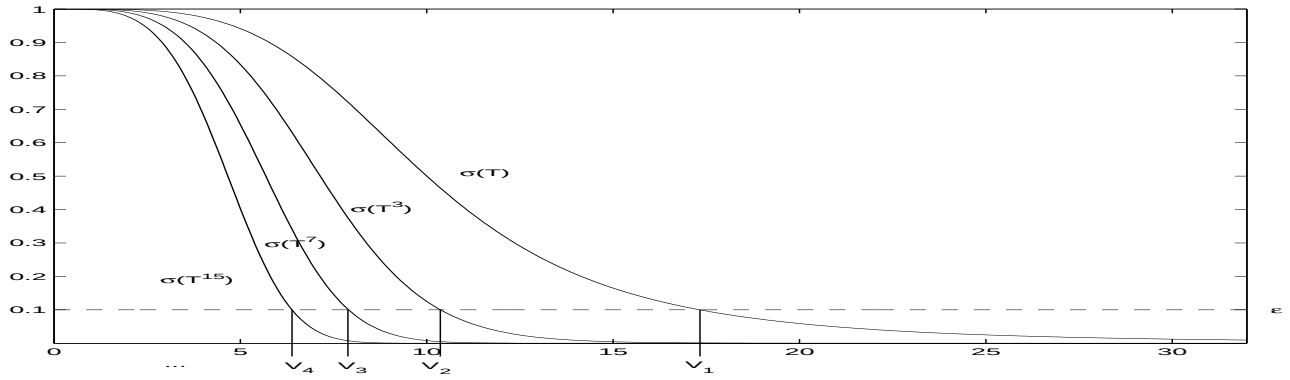

Figure 1. Spectra of powers of T and corresponding multiscale eigenspace decomposition. One can think of the horizontal axis as either indexing the eigenvalues of T and its powers, or indexing the corresponding eigenspaces.

geometrically on the space (such as Euclidean dilations, or rotations, or more general geometric transformations), the focus is on "dilation" operators acting on functions on the space. The key ingredient becomes a diffusion semigroup $\{T^t\}_{t\geq 0}$ with generator T, for example the heat semigroup on a manifold or graph (see the companion paper for the relevant standard definitions[7]). The operator T and its powers act on functions f by diffusing them. When $\{T^t\}$ is the heat semigroup, one can think of $T^t f$ as the distribution of heat at time t, starting with the initial heat distribution f at time $t = 0$. In many examples of great interest the spectrum of T is decaying. Hence for any pre-specified precision ϵ, the effective rank of T^t decreases with t, see Figure 1. We define the multiresolution subspaces V_j as the effective rank, up to precision ϵ, of T^{2^j-1}. This corresponds to a dyadic sampling of scales.

Given an initial highly localized orthonormal basis Φ_0 for the whole space (e.g., in the discrete case, Dirac δ-functions at each point of the space), we let V_j be the span of $T^{2^j-1}\Phi_0$. From the frequency point of view, V_j is approximately the span of the eigenfunctions corresponding to the eigenvalues λ such that $\lambda^{2^j-1} \geq \epsilon$, which are a set of "low-frequency" eigenfunctions, at least when smaller eigenvalues correspond to higher frequency eigenfunctions. We think of the set of functions $T^{2^j-1}\Phi_0$ as a set of smooth bump functions at scale j. These functions are in general highly overlapping. They are like the scaling functions in a wavelet pyramid with no downsampling. We enforce downsampling by constructing an orthonormal basis Φ_j, whose elements are (square-integrable) linear combinations of bumps $T^{2^j-1}\Phi_0$, and whose span, up to precision ϵ, is V_j. The cardinality of this orthonormal basis is equal to the ϵ-rank of T^{2^j-1}, and hence in general it is much smaller than the cardinality of the set of bumps $T^{2^j-1}\Phi_0$. This construction can be performed in a multiscale fashion by representing, at scale j, T^{2^j} onto Φ_j, considering the set of bumps $T^{2^j}\Phi_j$ and downsampling and orthonormalizing them to obtain Φ_{j+1}.

In Section 2 we review the original construction of orthogonal diffusion wavelets.

In Section 3 we present the generalization of this construction to the biorthogonal case.

In Section 4 we discuss three examples illustrating the new construction.

2. ORTHOGONAL DIFFUSION WAVELETS

We refer the reader to the companion paper[7] for relevant background and standard definitions of the Laplacian on a manifold or graph, and the corresponding heat kernel, as well as a summary of the original construction of orthogonal diffusion wavelets There are three key ingredients in the construction of orthonormal diffusion wavelets.[1] 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 critical 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.

In this section we present a particular case of our construction in a finite, discrete setting, for which only finite dimensional linear algebra is needed, and refer to the paper for the general construction.[1]

2.1. Setting and Assumptions

We consider a finite weighted graph[7, 15] X (which could be discretization of a compact manifold) and a symmetric positive definite and positive "diffusion" operator T on (functions on) X. In this paper we will consider the case $T = e^{-\epsilon\mathcal{L}}$, or $T = I - \epsilon\mathcal{L}$, where ϵ is a parameter, small and positive.

We assume that T is compact and self-adjoint, with spectrum $\lambda_0 = 1 \geq \lambda_1 \geq \ldots \lambda_j \geq \ldots$ (that accumulates only at 0) and corresponding eigenvectors $\xi_0, \xi_1, \ldots, \xi_j, \ldots$. We remark that the assumption that T is self-adjoint is not necessary, but simplifies the presentation. We also assume that T is local in the sense that if f has small support, Tf also has small support containing the support of f. Here and in all that follows, support should be understood as "numerical support". Moreover we assume that T is smoothing, and hence the numerical rank of powers of T decreases: see Figure 1. Ideally there exists a $\gamma < 1$ such that for every $j \geq 0$ we have $\mathrm{Ran}_\epsilon(T^{2^j}) < \gamma\,\mathrm{Ran}_\epsilon(T^{2^{j-1}})$, where Ran_ϵ denotes the ϵ-numerical rank,[1] i.e. the number of singular values of $T^{2^{j-1}}$ above ϵ.

We will compute and describe efficiently the powers T^{2^j}, for $j > 0$, which encode the long term behavior of the diffusion. We can interpret them as dilations acting on functions. We will define a multiresolution analysis associated with these dilations, and show how to construct orthonormal bases for the subspaces in the multiresolution analysis.

Some notation: here and in the rest of the paper we will use the notation $[L]_{B_1}^{B_2}$ to indicate the matrix representing the linear operator L with respect to the basis B_1 in the domain and B_2 in the range. A set of vectors B_1 represented on a basis B_2 will be written in matrix form $[B_1]_{B_2}$ or $[I]_{B_1}^{B_2}$ (with slight abuse of notation), where the columns of $[B_1]_{B_2}$ are the coordinates of the vectors B_1 in the coordinates B_2. We are not requiring that B_1 and B_2 span the same subspace. The notation is well-defined when $\langle B_1 \rangle \subseteq \langle B_2 \rangle$; when this is not the case we will explicitly define the meaning on a case-by-case basis.

2.2. Definition of the Multiresolution

By interpreting T as a dilation acting on functions in $L^2(X, \mu)$, we will define a natural multiresolution structure on X. First, we discretize the semigroup by letting $t_j = 2^j - 1$ and defining

$$\sigma_j = \{\lambda \in \sigma(T) : \lambda^{t_j} \geq \epsilon\}.$$

for a given precision ϵ. The set σ_j is the part of the spectrum of T^{t_j} above the precision ϵ; in essence, it is the "low-pass" portion of the spectrum. We now define subspaces V_j by:

$$V_j = \mathrm{span}(\{\xi_j : \lambda_j \in \sigma_j(T)\}).$$

Clearly, the sets V_j satisfy the basic properties of a multiresolution analysis:

1. $\cdots \subset V_j \subset V_{j-1} \ldots \subset V_1 \subset V_0 = L^2(X)$
2. $\lim_{j\to\infty} V_j = \mathrm{span}(\{\xi_i : \lambda_i = 1\})$
3. $\{\xi_j : \lambda_j \in \sigma_j\}$ is an orthonormal basis for V_j

The spaces V_j are adapted to the diffusion semigroup $\{T^t\}_t$. Essentially V_j is spanned by the low frequency eigenfunctions of T^{t_j}, where low-frequency simply means eigenvalue close to 1. In fact, in the case of the heat operator, this does correspond to a natural notion of frequency.[7] We can also interpret V_j as the significant low-pass portion of the spectrum of T^{t_j}, in the sense that T^{t_j} applied to any eigenvector outside of V_j yields a result "below precision."

We define the wavelet spaces W_j, $j \geq 0$, as the orthogonal complement of V_{j+1} in V_j so that

$$V_j = V_{j+1} \oplus^\perp W_j$$

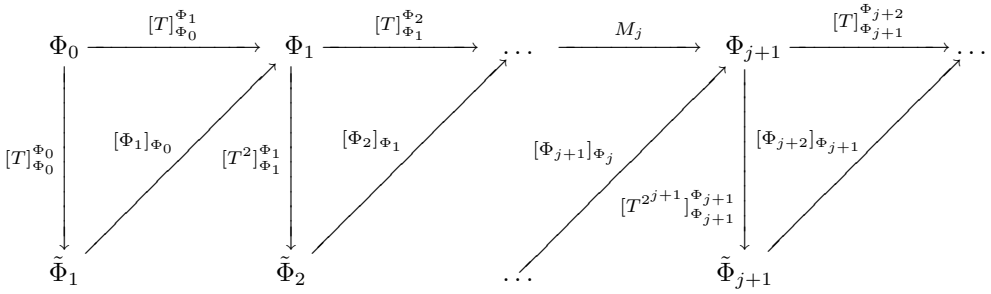

Figure 2. Diagram for downsampling, orthogonalization and operator compression. (All triangles are commutative by construction)

The subspaces V_j have decreasing (finite) dimension, depending on the decay of the spectrum of T. Ideally, we will have

$$\dim(V_j) \leq C_\epsilon 2^{-\alpha j}$$

where C_ϵ is a constant depending on the precision ϵ. The bases of eigenfunctions for the spaces V_j are, in general, highly nonlocal. The goal of both the orthogonal diffusion wavelet and biorthogonal diffusion wavelet constructions is to build localized smooth bases for the subspaces V_j.

2.3. Construction of the Multiresolution

We start by fixing a precision $\epsilon > 0$; we assume that T is self-adjoint and is represented on the basis $\Phi_0 = \{\delta_k\}_{k\in X}$, and consider the columns of T, which can be interpreted as the set of functions $\tilde{\Phi}_1 = \{T\delta_k\}_{k\in X}$ on X. We refer the reader to the diagram in Figure 2 for a visual representation of the scheme presented. We use a local multiscale orthogonalization procedure[1] to orthonormalize these columns and obtain a basis $\Phi_1 = \{\varphi_{1,k}\}_{k\in X_1}$ (X_1 is *defined* as this index set), written with respect to the basis Φ_0. This basis spans the range of T, which we denoted by V_1, up to precision ϵ, and is stored in the sparse matrix $[\Phi_1]_{\Phi_0}$. Essentially Φ_1 is a basis for the subspace V_1 which is ϵ-close to the range of T, and with basis elements that are well-localized. Moreover, the elements of Φ_1 are coarser than the elements of Φ_0, since they are the result of applying the "dilation" T once. Clearly $|X_1| \leq |X|$, but this inequality may already be strict since the numerical range of T may be approximated, within the specified precision ϵ, by a subspace of smaller dimension. Whether this is the case or not, we have the sparse matrix $[T]^{\Phi_1}_{\Phi_0}$, representing an ϵ-approximation of T with respect to Φ_0 in the domain and Φ_1 in the range. We can also represent T in the basis Φ_1: we denote this matrix by $[T]^{\Phi_1}_{\Phi_1}$. We can square this operator to obtain the dilation bringing us to the next scale: $[T^2]^{\Phi_1}_{\Phi_1} = [\Phi_1]_{\Phi_0}[T^2]^{\Phi_0}_{\Phi_0}[\Phi_1]^T_{\Phi_0} = [T]^{\Phi_1}_{\Phi_0}([T]^{\Phi_1}_{\Phi_0})^*$ (the last equality holds only when T is self-adjoint, and it is the only place where we use self-adjointness).

We now proceed by looking at the columns of $[T^2]^{\Phi_1}_{\Phi_1}$, which are $\tilde{\Phi}_2 = \{[T^2]^{\Phi_1}_{\Phi_1}\delta_k\}_{k\in X_1}$ i.e. $\{T^2\varphi_{1,k}\}_{k\in X_1}$ up to the precision ϵ. Applying a local orthonormalization procedure yields an orthonormal basis $\Phi_2 = \{\varphi_{2,k}\}_{k\in X_2}$ for the range of T_1^2 up to precision ϵ. Observe that Φ_2 is naturally written with respect to the basis Φ_1, and hence encoded in the matrix $[\Phi_2]_{\Phi_1}$, which plays the role of a low-pass filter, very much like the coefficients of a coarse scaling function onto the basis of scaling functions at the previous scale do. Moreover, depending on the decay of the spectrum of T, $|X_2|$ is in general a fraction of $|X_1|$. The matrix $[T^2]^{\Phi_2}_{\Phi_1}$ is then of size $|X_2| \times |X_1|$, and the matrix $[T^4]^{\Phi_2}_{\Phi_2} = [T^2]^{\Phi_2}_{\Phi_1}([T^2]^{\Phi_2}_{\Phi_1})^*$, a representation of T^4 acting on Φ_2, is of size $|X_2| \times |X_2|$.

After j steps in this fashion, we will have a representation of T^{2^j} onto a basis $\Phi_j = \{\varphi_{j,k}\}_{k\in X_j}$, encoded in a matrix $T_j := [T^{2^j}]^{\Phi_j}_{\Phi_j}$. The orthonormal basis Φ_j is represented with respect to Φ_{j-1}, and encoded in the matrix $[\Phi_j]_{\Phi_{j-1}}$. We let $\tilde{\Phi}_j = T_j\Phi_j$ We can represent the next dyadic power of T on Φ_{j+1} on the range of T^{2^j}. Depending on the decay of the spectrum of T, we expect $|X_j| << |X|$, in fact in the ideal situation the spectrum of T decays fast enough so that there exists $\gamma < 1$ such that $|X_j| < \gamma|X_{j-1}| < \cdots < \gamma^j|X|$. This corresponds to downsampling the set of columns of dyadic powers of T, thought of as vectors in $\mathcal{L}^2(X)$. The hypothesis that the rank of powers of T decreases guarantees that we can downsample and obtain coarser and coarser lattices in this spaces of columns.

While Φ_j is naturally identified with the set of Dirac δ-functions on X_j, we can extend these functions living on the "compressed" (or "downsampled") graph X_j to the whole initial graph X by writing

$$[\Phi_j]_{\Phi_0} = [\Phi_j]_{\Phi_{j-1}}[\Phi_{j-1}]_{\Phi_{j-2}} = \cdots = [\Phi_j]_{\Phi_{j-1}}[\Phi_{j-1}]_{\Phi_{j-2}} \cdot \cdots \cdot [\Phi_1]_{\Phi_0}[\Phi_0]_{\Phi_0} \,. \tag{1}$$

Since every function in Φ_0 is defined on X, so is every function in Φ_j. Hence any function on the compressed space X_j can be extended naturally to the whole X. In particular, one can compute low-frequency eigenfunctions on X_j in compressed form, and then extend them to the whole X. The elements in Φ_j are at scale $T^{2^{j+1}-1}$, and are much coarser and "smoother", than the initial elements in Φ_0, which is way they can be represented in compressed form. The projection of a function onto the subspace spanned by Φ_j will be by definition an approximation to that function at that particular scale.

Orthonormal bases of wavelets spanning the orthogonal complement of V_{j+1} into V_j can also be constructed.

One can easily compute the scaling function and wavelet transform in a way that is completely analogous to the classical Euclidean transforms. Suppose we are given f on X and want to compute $\langle f, \varphi_{j,k}\rangle$ for all scales j and corresponding "translations" k. Being given f means we are given $(\langle f, \varphi_{0,k}\rangle)_{k\in X}$. Then we can compute $(\langle f, \varphi_{1,k}\rangle)_{k\in X_1} = [\Phi_1]_{\Phi_0}(\langle f, \varphi_{0,k}\rangle)_{k\in X}$, and so on for all scales. The sparser the matrices $[\Phi_j]_{\Phi_{j-1}}$ (and $[T]_{\Phi_j}^{\Phi_j}$), the faster this computation. This generalizes the classical scaling function transform. We will show later that wavelets can be constructed as well, and that a fast wavelet transform is also possible.

As discussed in the paper,[1] these scaling function and wavelet bases have several of the typical properties that wavelets have in the Euclidean setting. They are well-localized in space, as well as in frequency. This means that the expansion of a wavelet in terms of the eigenfunctions of the operator T is well-localized at the corresponding scale. As a consequence they have vanishing moments, in the sense that wavelets at each scale are orthogonal to several low-frequency eigenfunctions. Also, having a compressed representation of dyadic powers of T, allow the efficient and precise computation of $T^k f$, for a wide range of k, both small, medium and large. In turn this allows the efficient computation of functions of the operator T, in particular the associated Green's function $(I - T)^{-1}$, as explained in the paper.[1]

3. BIORTHOGONAL DIFFUSION WAVELETS

In the classical setting, biorthogonal wavelet bases were introduced by Cohen, Daubechies, and Faveau.[16] By dropping the requirement that wavelet and scaling bases be orthonormal, they allow for more flexibility in wavelet design; for instance, while there are no smooth, orthogonal, compactly supported, symmetric wavelets, it is possible to construct compactly supported, smooth, symmetric biorthogonal wavelets. Biorthogonal wavelets have since found successful applications in several areas of signal and image processing.

The diffusion wavelet construction builds smooth, local orthonormal bases for the scaling and wavelet spaces V_j and W_j. By generalizing the construction to allow for biorthogonal bases, as in the classical setting, one introduces an extra degree of flexibility which might be exploited.

Of particular interest, is the possibility of constructing "sparser" bases; i.e., bases whose elements have smaller support. After all, one of the primary motivations for diffusion wavelets is the desire to build bases well adapted to the spectrum of a diffusion operator, but more compactly supported than bases consisting of eigenvectors. The diffusion basis for the space V_j spans approximately the same space as the eigenvectors $\{e_k \mid \lambda_k^{2^j} \geq \epsilon\}$, but the diffusion wavelets are concentrated on a relatively small set with exponential decay whereas the eigenvectors e_k are (in typical cases) supported on the entire graph. The orthonormal bases generated by the diffusion wavelet construction are built from sums of selected columns of the input matrices T_j by the orthogonalization algorithm. As such, they span the same subspace as a set of columns of T_j, but they are less compactly supported. This strongly suggests simply choosing a set of columns of the input matrix T_j as an biorthogonal basis for the scaling space. This offers other advantages as well; in the case of Markov chains it is convenient to represents the states at a certain time scale in terms of probability distributions at the same scale, and the columns of the corresponding power of the Markov matrix are natural candidates.

3.1. Biorthogonal Multiresolution Analysis

In the biorthogonal case there are two dual multiresolution analysis. The primal multiresolution analysis $\{V_j\}_{j\geq 0}$ is defined as in the orthogonal case, except that a Riesz basis Φ_j spanning V_j is constructed, instead of the an orthonormal basis. Then a dual multiresolution analysis $\{\tilde{V}_j\}_{j\geq 0}$ is constructed in such a way that $\tilde{V}_j$ is spanned by a basis $\tilde{\Phi}_j$ of scaling functions which are a dual Riesz basis to Φ_j, i.e. $\langle \varphi_{j,k}, \tilde{\varphi}_{j,k'} \rangle = \delta_{k,k'}$ for any $k \neq k' \in \mathcal{K}_j$. Hence for any function f in V_j we have

$$f = \sum_{k\in\mathcal{K}_j} \langle f, \tilde{\varphi}_{j,k} \rangle \varphi_{j,k}$$

and furthermore this decomposition is stable, in the sense that there exist constants $A, B > 0$ independent of f such that

$$A||f||_2 \leq \sum |\langle f, \tilde{\varphi}_{j,k} \rangle|^2 \leq B||f||_2 \,.$$

In general the operator

$$f \mapsto \sum_{k\in\mathcal{K}_j} \langle f, \tilde{\varphi}_{j,k} \rangle \varphi_{j,k}$$

is an oblique projection on V_j along $\tilde{V}_j$, and analogously

$$f \mapsto \sum_{k\in\mathcal{K}_j} \langle f, \varphi_{j,k} \rangle \tilde{\varphi}_{j,k}$$

is an oblique projection on $\tilde{V}_j$ along V_j.

There are also two families of wavelet subspaces, $\{W_j\}$ and $\{\tilde{W}_j\}$, with the property that $V_j = V_{j+1} \oplus W_j$ and $\tilde{V}_j = \tilde{V}_{j+1} \oplus \tilde{W}_j$, where the sum is direct but not necessarily orthogonal. On the other hand it is always the case that W_j is orthogonal to $\tilde{V}_{j+1}$ and $\tilde{W}_j$ is orthogonal to V_{j+1}. In general for a fixed multiresolution $\{V_j\}$ there are several dual multiresolutions $\{\tilde{V}_j\}$ with the properties above.

3.2. Multiscale Biorthogonal Construction

The construction is designed to take advantage of the fact that the operators T^{2^j} are increasingly rank deficient. This means that the effective numerical dimension of the space V_j, which comprises the part of the spectrum of T^{2^j} above precision, is decreasing in j. Since V_{j+1} is a subset of V_j, we can always represent objects at the $j+1$ level of the multiresolution analysis in terms of V_j, the basis at the preceding level. This representation means that all the objects we store and operator on are effectively compressed.

We start from a basis Φ_0 which is ϵ-dense in V_0 and a basis $\tilde{\Phi_0}$ which is ϵ-dense is $\tilde{V_0}$ and dual to Φ_0;

$$\tilde{\Phi_0}^* \Phi_0 = I$$

We are also given a diffusion operator T represented with respect to the basis Φ_0; that is, we are given the matrix $T_0 = [T]_{\Phi_0}^{\Phi_0}$. We will assume that T_0 is sparse.

We use one of the orthogonalization procedures discussed in the preceding section to approximately factor the $n \times n$ matrix T_0 as

$$T_0 \Pi = \begin{pmatrix} T_{11} & \vert & T_{12} \end{pmatrix} = Q \begin{pmatrix} R_{11} & \vert & R_{12} \end{pmatrix}$$

where Π is an $n \times n$ permutation, T_{11} is a $n \times k$ matrix consisting of k columns of T_0, T_{12} is a $n \times (n-k)$ matrix consisting of the remaining columns of T_0, Q is a $n \times k$ orthogonal matrix, R_{11} is a $k \times k$ upper triangular matrix and R_{12} is an $k \times (n-k)$ matrix. We interpret this factorization as follows. The columns of the matrix Q form a downsampled orthogonal basis Θ_1 which spans the range space of T_0 up to precision. The columns of Q are formed from sums of the columns of T_{11}, which are those columns of the input matrix T_0 "chosen" by the orthogonalization procedure. The matrix R_{11} encodes the information needed to form T_{11} from the orthogonal basis Θ_1; the matrix R_{12} is used to approximate the remaining columns, up to the specified precision, using the basis Θ_1. The matrix $\begin{pmatrix} R_{11} & \vert & R_{12} \end{pmatrix}$ is $[T]_{\Phi_0}^{\Theta_1}$, which represents the operator T on the domain Φ_0 with range in Θ_1.

In the biorthogonal construction, we discard the actual orthogonal basis Θ_1 and define the filter matrix M_1 by $M_1 = T_{11} = QR_{11}$, which is the set of columns of T_0 "chosen" by the orthogonalization procedure. Of course, these vectors span the same set as the orthogonal basis Θ_1. We define the next basis Φ_1 as follows:

$$\Phi_1 = \Phi_0 M_1 = \Phi_0 T_{11} = \Phi_0 Q R_{11}$$

Of course we only store the matrix M_1, which is a representation $[I]_{\Phi_1}^{\Phi_0}$ of the basis Φ_1 in terms of the initial basis Φ_0.

Since the basis Φ_1 is not orthogonal, we will need to construct a dual basis $\tilde{\Phi}_1$. Recalling that $T_{11} = QR_{11}$, one might naively decide to set

$$\tilde{\Phi}_1 = \tilde{\Phi}_0 Q R_{11}^{-*}. \tag{2}$$

This is, in fact, a dual basis since

$$\tilde{\Phi_1}^* \Phi_1 = R_{11}^{-1} Q^* \tilde{\Phi_0}^* \Phi_0 Q R_{11} = I.$$

There is, however, a serious problem with this simplistic approach. The matrix Q in (2) implicitly includes the downsampling operator which was applied to the columns of T_0. But there is no reason to believe that the columns of $\tilde{\Phi}_1$ should be downsampled in the same manner as T_0. In general, the downsampling operator applied to $\tilde{\Phi}_1$ will not choose a basis which spans $\tilde{V}_1$ to the proper precision. A proper dual basis can be formed by taking $\tilde{\Phi}_1$ to be a pseudoinverse of the matrix Φ_1. A filter matrix $\tilde{M}_1$ such that

$$\tilde{\Phi}_1 = \tilde{\Phi}_0 \tilde{M}_1$$

can then be computed. Of course, this is not an efficient method for computing the dual basis. The authors are presently investigating how to efficiently construct dual bases in a natural multiscale fashion.

The matrix $\tilde{M}_1$ is a representation $[I]_{\tilde{\Phi}_1}^{\tilde{\Phi}_0}$ of the basis $\tilde{\Phi}_1$ with respect to $\tilde{\Phi}_0$; but there is another interpretation as well. The coefficients of a vector in the basis Φ_1 are computed by taking inner products of that vector with the elements of the dual basis $\tilde{\Phi}_1$. If the individual elements of Φ_1 are denoted by $\{\phi_1^j\}_j$ and those of $\tilde{\Phi}_1$ are denote by $\{\tilde{\phi}_1^j\}_j$ then for a vector x in V_1

$$x = \sum_j < x, \tilde{\phi}_1^j > \phi_1^j$$

This implies that $\tilde{M_1}^*$ is the change of basis matrix $[I]_{\Phi_0}^{\Phi_1}$ which takes vectors represented in Φ_0 and computes their coefficients in Φ_1.

We next form a representation of the operator T^2 on the basis Φ_1. In order to do this, we first need to compute a representation of T on the domain Φ_0 with range Φ_1. This matrix, which we will call Proj_1, is given by

$$\begin{pmatrix} I & R_{11}^{-1} R_{12} \end{pmatrix} \Pi^*$$

To see that this is, in fact, correct note that

$$\begin{aligned} \Phi_1 \mathrm{Proj}_1 &= \Phi_0 Q_{11} R_{11} \begin{pmatrix} I & R_{11}^{-1} R_{12} \end{pmatrix} \Pi^* \\ &\approx \begin{pmatrix} T_{11} & T_{12} \end{pmatrix} \Pi^* \\ &= T_0 \end{aligned}$$

The square is then given by

$$\begin{aligned} [T^2]_{\Phi_1}^{\Phi_1} &= [I]_{\Phi_0}^{\Phi_1} [T]_{\Phi_0}^{\Phi_0} [T]_{\Phi_1}^{\Phi_0} \\ &= \tilde{M_1}^* T_0 * \mathrm{Proj}_1 \end{aligned}$$

The resulting matrix, which we will call T_1, becomes the input to the next level of the algorithm; we would now proceed by applying the orthogonalization procedure to the columns of T_1 in order to downsample them, and so on.

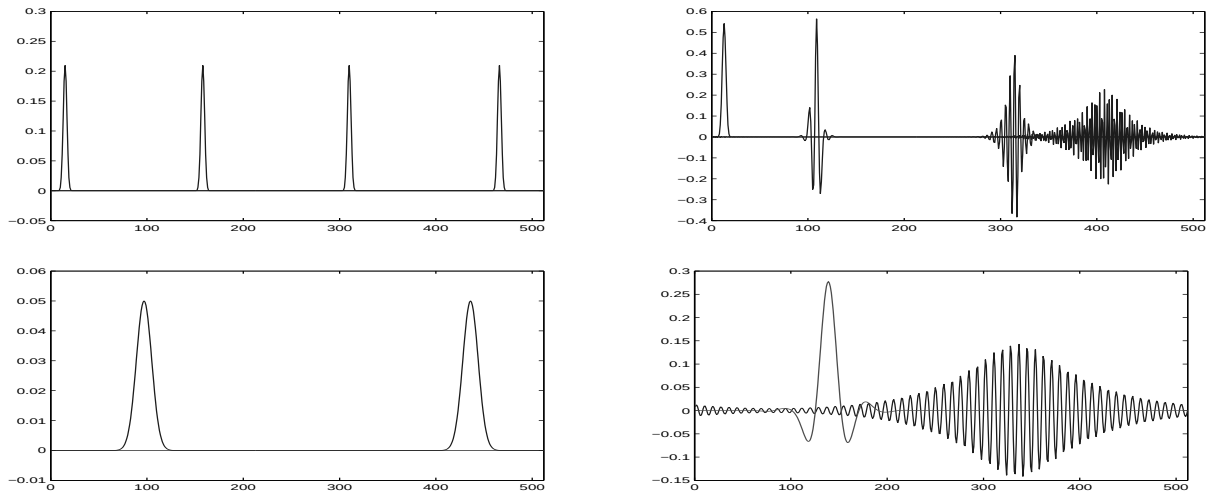

Figure 3. Some biorthogonal scaling functions in V_3 (top left), in V_7 (bottom left), and some orthogonal scaling functions in V_3 (top right), and V_7 (bottom right).

The end result of the algorithm is a cascading sequence of bases and dual bases, each represented on the bases at the preceding level via a filter matrix:

$$\begin{aligned} \Phi_0 \qquad & \tilde{\Phi}_0 \\ \Phi_1 = \Phi_0 M_1 \qquad & \tilde{\Phi}_1 = \tilde{\Phi}_0 \tilde{M}_1 \\ \Phi_2 = \Phi_1 M_2 \qquad & \tilde{\Phi}_1 = \tilde{\Phi}_1 \tilde{M}_2 \\ & \dots \end{aligned}$$

where Φ_j approximately spans V_j, $\tilde{\Phi}_j$ approximately spans $\tilde{V}_j$, and $\tilde{\Phi_j}^* \Phi_j$ is the identity. This generalizes classical biorthogonal wavelets in the sense that scaling functions and wavelets are encoded in terms of cascading filters. In the classical setting the filters are the same across levels and locations, because of the dilation and translation invariance of the construction. In this case no symmetries are assumed, so the filter matrices do not exhibit any obvious structure.

4. EXAMPLES

In this section, we look at three examples. We start by illustrating the biorthogonal construction in a particularly simple setting—the circle—and compare the resulting scaling functions to those obtained from the original diffusion wavelet construction. The second example is the construction of the multiresolution analysis on the unit sphere. In the final example, we consider a Markov process on a 8-state space. We will see that one of the advantages of the biorthogonal construction is that the bases are always columns of a power of the original operator (although those powers are not stored with respect to the original basis).

4.1. Multiresolution diffusion on the Circle

We consider a discretization of the Laplacian on the circle S^1. We let X be a uniformly sampled grid of 512 points on $[0, 2\pi]$ and T be the "standard" discretization of the Laplacian represented with respect to the delta basis on X:

We take both the initial basis Φ_0 and the initial dual basis $\tilde{\Phi}_0$ to be the delta basis on X, and set the precision, ϵ, to 10^{-5}. The biorthogonal construction proceeds by applying T to the basis Φ_0 and downsampling

$$T = \begin{pmatrix} 0.50 & 0.25 & \dots & \dots \\ 0.25 & 0.50 & 0.25 & \dots \\ \dots & 0.25 & 0.50 & \dots \\ \dots & \dots & 0.25 & \dots \\ \dots & \dots & \dots & \dots \\ 0.25 & \dots & \dots & \dots \end{pmatrix}$$

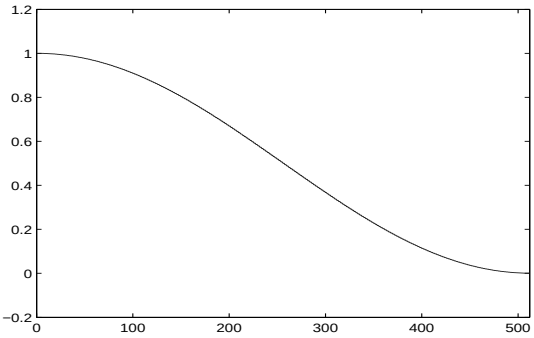

Figure 4. The spectrum of the diffusion operator T on the circle.

the columns of T to obtain a Riesz basis Φ_1 of functions for V_1. This basis is represented in terms of the basis Φ_0; that is, there is a filter matrix M_1 such that

$$\Phi_1 = \Phi_0 M_1$$

The downsampling is achieved by using the same orthogonalization process as in the orthogonal case, but the orthogonalized basis is discarded and the columns of T which were used to build that orthogonal basis are retained. Next, A dual basis $\tilde{\Phi}_1 = \tilde{\Phi}_0 \tilde{M}_1$ such that

$$\tilde{\Phi_1}^* \Phi_1 = I$$

is computed in terms of the dual basis $\tilde{\Phi}_0$ at the proceeding level. At the present time, the authors do not know how to do this dual basis computation in a multiscale fashion; however, the dual basis can always be computed by taking the pseudoinverse of Φ_1 in the original delta basis and changing the representation of the pseudoinverse. Next, a representation T_2 of the matrix T^2 with respect to the new basis Φ_1 is computed. This computation involves both the filter M_1 and the dual filter $\tilde{M}_1$. The process is now repeated; that is, the columns of the new matrix T_1 are downsampled to form a basis $\Phi_2 = \Phi_1 M_2$ and a basis $\tilde{\Phi}_2 = \tilde{\Phi}_1 \tilde{M}_2$ dual to Φ_2 is computed. The algorithm continues in this fashion, until the desired level is reached, in this case we stop at level 10.

This process differs from the orthogonal construction in two crucial respects. First, the orthogonalization process is used only to downsample the input matrix T_j, the actual orthogonal basis is discarded; and second, since the bases are no longer orthogonal, a dual basis must be computed. At each step in the orthogonal construction, the orthogonalization process is applied to the input operator T_j, yielding an orthogonal filter matrix N_j. The matrix N_j is constructed with a Gram-Schmidt process, so the resulting basis $\Phi_j = \Phi_{j-1} N_j$ consists of several layers. The first layer of the basis Φ_j corresponds to the first few columns of N_j, which are simply selected columns of the input matrix T_j. These functions have very small support. The orthogonalization

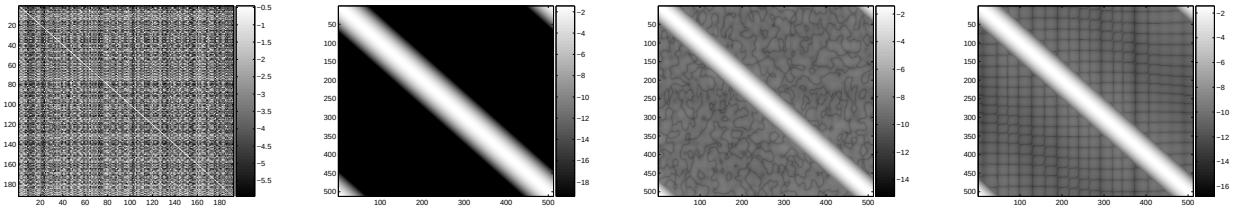

Figure 5. From left to right, all in logarithmic scale: compressed representation of the operator T^{128}, the true operator T^{128}. Notice that the matrix on the left is 192×192, while the one on the right is 512×512. Reconstruction of T^{128} from the compressed biorthogonal representation and reconstruction of T^{128} from the compressed orthogonal representation.

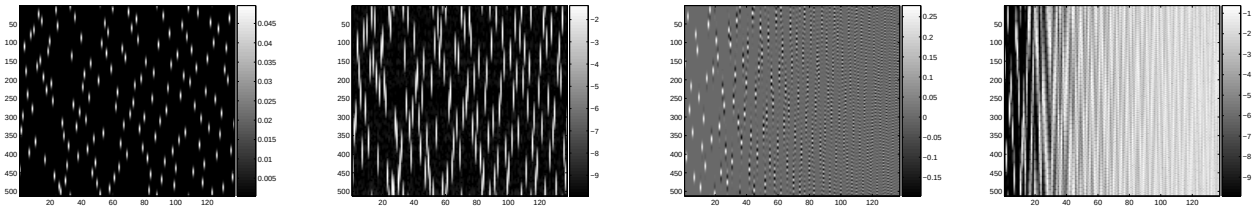

Figure 6. From left to right: uncompressed representations of the biorthogonal basis on the circle at level 7, same matrix in logarithmic scale; the (layered) orthogonal basis on the circle at level 7, same matrix in logarithmic scale.

process takes sums of a small number of columns of T_j in order to build the next few orthogonal columns of N_j; this corresponds to the second layer of Φ_j. These functions necessarily have larger support. The orthogonalization process downsamples the matrix T_j; it only uses a subset of the columns of T_j in order to construct N_j. But the last few columns of N_j consist of functions which are sums over the full set of downsampled columns of T_j. This corresponds to the last layer of functions in Φ_j, which generally have much larger support than those in the first few layers.

Figure 3 shows biorthogonal scaling functions at various levels and compares them to orthogonal scaling functions at the same levels. Notice that the biorthogonal scaling functions at each level all have supports of the same size, unlike the orthogonal bases, which are layered. The functions in the first layer of each orthogonal basis are formed from a single column of the input matrix, but the functions in later layers are sums over an ever increasing number of the columns of the input matrix and as a result their support grow.

Figure 5 shows the compressed representation of T^{128} with respect to Φ_8 and the "true" operator T^{128}. The compressed version is a 192×192 matrix with $36,864$ nonzero entries while the true T^{128} is a 512×512 matrix with $131,584$ entries. Figure 5 shows the reconstruction of T^{128} from its compressed form in both the orthogonal and biorthogonal case. The L^2 operator norm of the difference between the reconsutrction versions and the "true" operator T^{128} is smaller than the required precision, 10^{-5}, in both cases.

Finally, Figure 6 shows uncompressed representations of the orthogonal basis for V_7 and the biorthogonal scaling function basis for V_7. Here the matrices Φ_j are displayed, so each column represents a single basis function. The biorthogonal basis is simply a permuted subset of the columns of T^{128}, while the orthogonal basis shows the layering effect quite clearly.

4.2. Multiresolution Analysis on the Sphere

Diffusion wavelets would be of little use if they were restricted to the circle. For our next example, we will take X to be set of 2000 uniformly distributed points $p_1, p_2, \ldots, p_{2000}$ on the unit sphere S^2. The matrix for the diffusion operator T with respect to the delta basis on X is given by

$$T(i,j) = \exp(-\alpha \cdot ||p_i - p_j||_2^2)$$

where α is a constant. Entries that fall below the fixed threshold, 10^{-10}, are discarded. Assuming α is reasonable, this matrix is sparse and can be constructed efficiently by performing a nearest neighbor search for each point in X (the columns of T are just truncated bump functions).

Once again we run both the orthogonal and biorthogonal algorithms with precision set to 10^{-5}. Figure 8 compares some orthogonal and biorthogonal scaling functions at level 5 of the construction.

Figure 8 compares the compressed representation of T^8 to the "true" operator T^8. The uncompressed T^8 has more than $690,000$ entries with magnitude above 10^{-10}, while the compressed version has approximately $400,000$ entries above the same threshold.

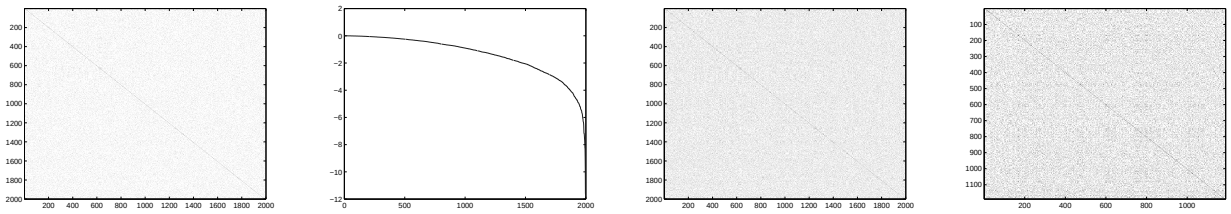

Figure 7. From left to right: plot of the nonzero entries of the operator T on the sphere above 10^{-6} and its spectrum, the matrix T^8 and the compressed version of T^8.

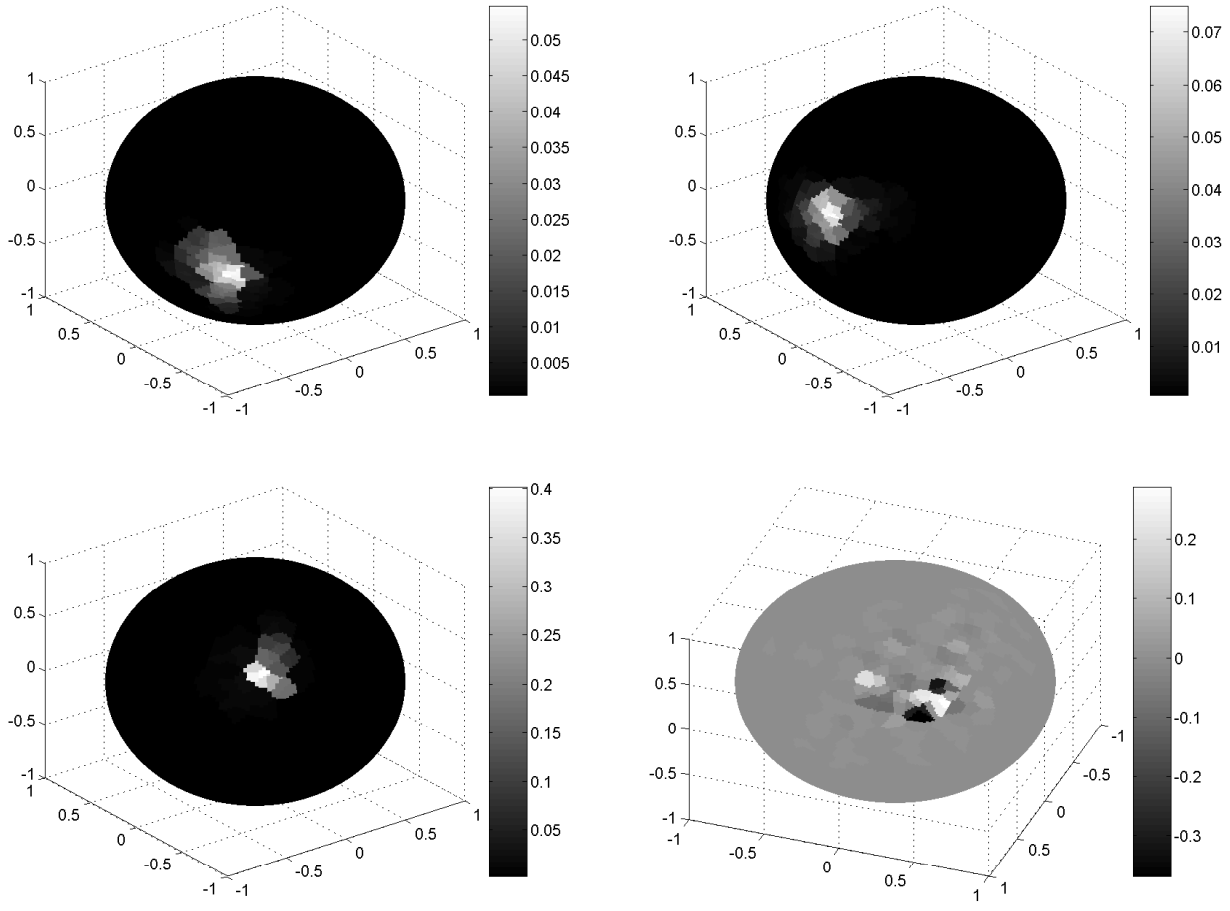

Figure 8. Biorthogonal scaling functions at level 5 (top left and top right), and some orthogonal scaling at the same level (bottom left and bottom right).

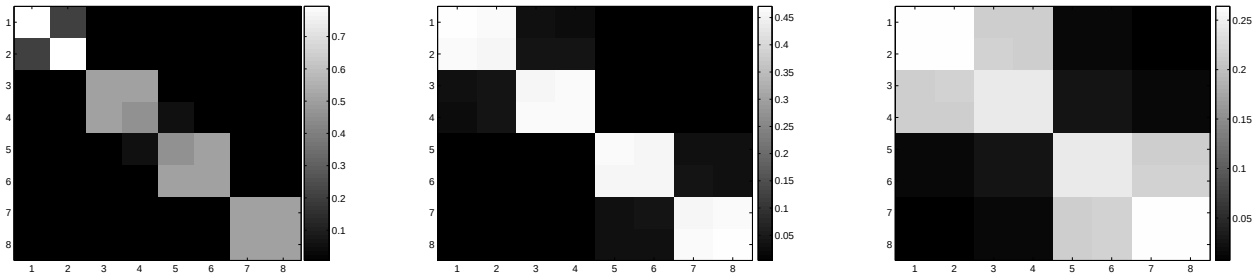

Figure 9. Some dyadic powers of the Markov chain T.

4.3. Markov Chain

We now consider a simple example of a Markov chain on a graph with 8 states. We label the states $\{\nu_1, \nu_2, \ldots, \nu_8\}$ and let $T(i,j)$ be the probability of moving from state ν_i to ν_j, where T is the matrix

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

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. Figure 9 shows some dyadic powers of the operator T.

Figure 10 shows compressed representations of T at several different levels. These compressed representations show the structure of the Markov chain clearly. The operator T_6 has downsampled the number of states down to 4, one for each of the pairs. One can see the structure of weak and strong interactions between these pairs from the matrix for T_6. By level 13 (Figure 10), downsampling leaves only two states, one corresponding to the original states $\{\nu_1, \ldots, \nu_4\}$ and one corresponding to $\{\nu_5, \ldots, \nu_8\}$.

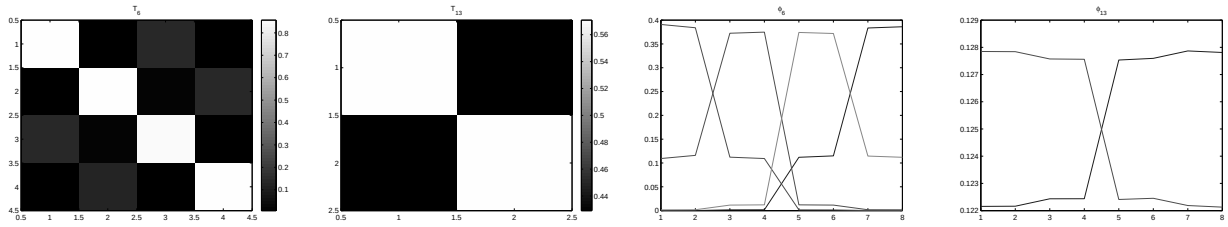

Figure 10. From left to right: compressed representation T_6 (left)and T_{13}, Level 6 (left) and level 13 scaling functions.

5. ACKNOWLEDGEMENTS

The first author is partially supported by NSF, under DMS Grant 0512050.

REFERENCES

1. R. Coifman and M. Maggioni, "Diffusion wavelets," *Tech. Rep. YALE/DCS/TR-1303, Yale Univ., Appl. Comp. Harm. Anal.* , Sep. 2004. to appear.

2. 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.
3. L. Greengard and V. Rokhlin, "A fast algorithm for particle simulations," *J Comput Phys* **73**, pp. 325–348, 1987.
4. G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet tranforms and numerical algorithms," *Comm Pure Applied math* **44**, pp. 141–183, 1991.
5. A. Brandt, "Algebraic multigrid theory: the symmetric case," *Appl. Math. Comp.* (19), pp. 23–56, 1986.
6. A. Nahmod, *Geometry of operators and spectral analysis*. PhD thesis, Yale University, Dept of Math., 1991. RR Coifman.
7. A. D. Szlam, M. Maggioni, R. R. Coifman, and J. C. B. Jr, "Diffusion-driven multiscale analysis on manifolds and graphs: top-down and bottom-up constructions," August 2005. SPIE Wavelet XI.
8. R. Coifman and S. Lafon, "Geometric harmonics," tech. rep., Yale University, Dept. Comp. Sci., 2003.
9. S. Lafon, *Diffusion maps and geometric harmonics*. PhD thesis, Yale University, Dept of Mathematics & Applied Mathematics, 2004.
10. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comp. Harm. Anal.* , 2004.
11. R. Coifman and S. Lafon, "Geometric harmonics," *Appl. Comp. Harm. Anal.* , 2004 2004. Submitted.
12. M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation* **6**, pp. 1373–1396, June 2003.
13. P. Niyogi and M. Belkin, "Semi-supervised learning on Riemannian manifolds," Tech. Rep. TR-2001-30, University of Chicago, Computer Science Dept, Nov. 2001.
14. E. Stein, *Topics in Harmonic Analysis related to the Littlewood-Paley theory*, Princeton University Press, 1970.
15. F. Chung, *Spectral Graph Theory*, no. 92, CBMS-AMS, May 1997.
16. A. Cohen, I. Daubechies, and J. Faveau, "Biorthogonal bases of compactly supported wavelets," *Comm Pure Applied Math* **XLV**, pp. 485–560, 1992.