MULTISCALE GEOMETRIC DICTIONARIES FOR POINT-CLOUD DATA

Guangliang Chen[1], *Mauro Maggioni*[1,2]

[1]Mathematics and [2]Computer Science Departments, Duke University, Durham, NC 27708, U.S.A.
Email: {glchen,mauro}@math.duke.edu

ABSTRACT

We develop a novel geometric multiresolution analysis for analyzing intrinsically low-dimensional point clouds in high-dimensional spaces, modeled as samples from a d-dimensional set $\mathcal{M}$ (in particular, a manifold) embedded in $\mathbb{R}^D$, in the regime $d \ll D$. This type of situation has been recognized as important in various applications, such as the analysis of sounds, images, and gene arrays. In this paper we construct data-dependent multiscale dictionaries that aim at efficient encoding and manipulating of the data. Unlike existing constructions, our construction is fast, and so are the algorithms that map data points to dictionary coefficients and vice versa. In addition, data points have a guaranteed sparsity in terms of the dictionary.

Keywords— Data Sets. Point Clouds. Wavelets. Dictionary Learning. Multiscale Analysis. Sparse Approximation.

1. INTRODUCTION

Data sets are often modeled as point clouds in $\mathbb{R}^D$, for D large, but having some interesting low-dimensional structure, for example that of a d-dimensional manifold $\mathcal{M}$, with $d \ll D$. When $\mathcal{M}$ is simply a linear subspace, one may exploit this assumption for encoding efficiently the data by projecting onto a dictionary of d vectors in $\mathbb{R}^D$ (for example found by SVD), at a cost $(n + D)d$ for n data points. When $\mathcal{M}$ is nonlinear, there are no "explicit" constructions of dictionaries that achieve a similar efficiency: typically one uses either random dictionaries, or dictionaries obtained by black-box optimization. This type of situation has been recognized as important in various applications, ranging from the analysis of sounds, images (RGB or hyperspectral), to gene arrays, EEG signals, and other types of manifold-valued data, and has been at the center of much investigation in the applied mathematics [4] and machine learning communities during the past several years.

We formalize this approach by requesting to find a dictionary Φ of size I, using training data, such that every point (at least from the training data set) may be represented, up to a certain precision ϵ, by at most m elements of the dictionary. This requirement of sparsity of the representation is very natural from the point of statistics, signal processing, and interpretation of the representation. Of course, the smaller I and m are, for a given ϵ, the better the dictionary.

Current constructions of these dictionaries such as K-SVD [1] , k-flats [8] and others have several deficiencies. First, they cast the requirements above as an optimization problem, with many local minima, and for iterative algorithms little is known about their computational complexity. Second, no guarantees are provided about the size of I, m (as a function of ϵ). Lastly, the dictionaries found are in general highly overcomplete and unstructured. As a consequence, there is no fast algorithm for computing the coefficients of the representation of a data point in the dictionary, thus requiring appropriate sparsity-seeking algorithms.

In this paper we construct data-dependent dictionaries based on a multiresolution geometric analysis of the data, inspired by multiscale techniques in geometric measure theory [5]. These dictionaries are structured in a multiscale fashion; they can be computed efficiently; the expansion of a data point on the dictionary elements is guaranteed to have a certain degree of sparsity m, and may be computed by a fast algorithm; the growth of the number of dictionary elements I as a function of ϵ is controlled theoretically, and easy to estimate in practice. We call the elements of these dictionaries *geometric wavelets*, since in some aspects they generalize wavelets from vectors that analyze functions to affine vectors that analyze point clouds.

2. MULTISCALE GEOMETRIC WAVELETS

Let $(\mathcal{M}, g, \mu)$ be a compact Riemannian manifold of dimension d isometrically embedded in $\mathbb{R}^D$ (with $d \ll D$) and endowed with the natural volume measure. Our construction consists of three steps: First, we perform a nested geometric decomposition of $\mathcal{M}$ into dyadic cubes at a total of J scales, arranged in a tree. Second, we obtain a d-dimensional affine approximation in each cube, generating a sequence of piecewise linear sets $\mathcal{M}_j, j \leq J$ approximating the manifold. Lastly, we construct low-dimensional affine difference operators that efficiently encode the differences between $\mathcal{M}_j$ and $\mathcal{M}_{j+1}$.

This construction parallels, in a geometric setting, that of classical multiscale wavelet analysis, and we therefore call it a Geometric Wavelets MultiResolution Analysis (GWMRA). We show that when $\mathcal{M}$ is a smooth manifold, guarantees on the approximation rates of $\mathcal{M}$ in terms of the $\mathcal{M}_j$ are easily derived. We construct bases for the various affine operators involved,

The authors thank E. Monson for useful discussions. MM is grateful for partial support from NSF, ONR, and the Sloan Foundation. GC was partially supported by ONR N00014-07-1-0625 and NSF CCF 0808847.

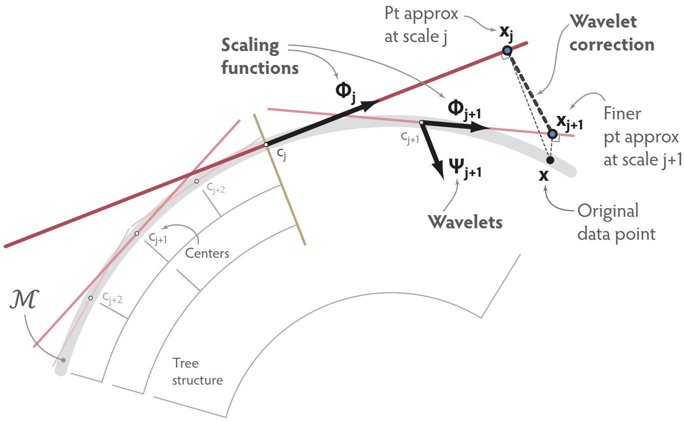

Fig. 1. Construction of geometric wavelets (figure by E. Monson).

producing a hierarchically organized dictionary that is adapted to the data, which we expect to be useful in many applications. Finally, efficient algorithms exist for computing both the multiscale structure above and the geometric wavelet transform and its inverse. Such efficient algorithms are currently not available for any of the algorithms in the dictionary-learning community.

Multiscale cubes. For any $x \in \mathcal{M}$ and $r > 0$, we use $B_r(x)$ to denote the ball in the manifold $\mathcal{M}$ of radius r centered at x. We start by a spatial multiscale decomposition of $\mathcal{M}$ into dyadic cubes, $\{C_{j,k}\}_{k\in\Gamma_j, j\in\mathbb{Z}}$, which are open sets in $\mathcal{M}$ such that

(i) for every $j \in \mathbb{Z}$, $\mu(\mathcal{M} \setminus \cup_{k\in\Gamma_j} C_{j,k}) = 0$;

(ii) for $j' \geq j$ either $C_{j',k'} \subseteq C_{j,k}$ or $\mu(C_{j,k} \cap C_{j',k'}) = 0$;

(iii) for any $j < j'$ and $k' \in \Gamma_{j'}$, there exists a unique $k \in \Gamma_j$ such that $C_{j',k'} \subseteq C_{j,k}$;

(iv) each $C_{j,k}$ contains a point $c_{j,k}$, called center of $C_{j,k}$, such that $B_{c_1\cdot 2^{-j}}(c_{j,k}) \subseteq C_{j,k} \subseteq B_{\min\{c_2\cdot 2^{-j},\mathrm{diam}(\mathcal{M})\}}(c_{j,k})$, for fixed constants c_1, c_2 depending on intrinsic geometric properties of $\mathcal{M}$. In particular, we have $\mu(C_{j,k}) \sim 2^{-dj}$;

(v) the boundary of each $C_{j,k}$ is piecewise smooth.

The construction of such dyadic cubes is possible on rather general spaces [5]. The family of dyadic cubes at scale j generates a σ-algebra $\mathcal{F}_j$. Functions measurable with respect to this σ-algebra are piecewise constant on each cube.

The properties above imply that there is a natural tree structure $\mathcal{T}$ associated to the family of dyadic cubes: for any (j, k), we let $\mathrm{children}(j, k) = \{k' \in \Gamma_{j+1} : C_{j+1,k'} \subseteq C_{j,k}\}$. Note that $C_{j,k}$ is the disjoint union of its children. For every $x \in \mathcal{M}$, with abuse of notation we let (j, x) be the unique $k(x) \in \Gamma_j$ such that $x \in C_{j,k(x)}$. We assume $\mu(\mathcal{M}) \sim 1$ so that there is only one dyadic cube at the root of the tree (thus we will only consider $j \geq 0$).

Multiscale singular value decompositions (MSVD). The tools we build upon are classical in multiscale geometric measure theory [5], especially in its intersection with harmonic analysis. An introduction to these with an application to estimation of intrinsic dimension of point clouds is in [7] and references therein.

We start with some geometric objects that are associated to the dyadic cubes. For each $C_{j,k}$ we define the mean

$$c_{j,k} := \mathbb{E}_\mu[x|x \in C_{j,x}] = \frac{1}{\mu(C_{j,k})}\int_{C_{j,k}} x\, d\mu(x), \tag{2.1}$$

and the covariance operator restricted to $C_{j,k}$,

$$\mathrm{cov}_{j,k} = \mathbb{E}_\mu[(x - c_{j,k})(x - c_{j,k})^*|x \in C_{j,k}]. \tag{2.2}$$

Let the rank-d Singular Value Decomposition (SVD) be

$$\mathrm{cov}_{j,k} \approx \Phi_{j,k}\Sigma_{j,k}\Phi_{j,k}^*, \tag{2.3}$$

and define the approximate local tangent space

$$\mathbb{V}_{j,k} := V_{j,k} + c_{j,k}, \quad V_{j,k} = \langle \Phi_{j,k} \rangle\,, \tag{2.4}$$

where $\langle \Phi_{j,k} \rangle$ denotes the span of the columns of $\Phi_{j,k}$, so that $\mathbb{V}_{j,k}$ is the affine subspace of dimension d parallel to $V_{j,k}$ and passing through $c_{j,k}$. We think of $\{\Phi_{j,k}\}_{k\in\Gamma_j}$ as the geometric analogue of scaling functions at scale j. Let $\mathbb{P}_{j,k}$ be the associated affine projection onto $\mathbb{V}_{j,k}$: for any $x \in C_{j,k}$,

$$\mathbb{P}_{j,k}(x) := P_{j,k} \cdot (x - c_{j,k}) + c_{j,k}, \quad P_{j,k} = \Phi_{j,k}\Phi_{j,k}^*, \tag{2.5}$$

and define a coarse approximation of $\mathcal{M}$ at scale j,

$$\mathcal{M}_j := \cup_{k\in\Gamma_j}\mathbb{P}_{j,k}(C_{j,k}), \tag{2.6}$$

which is the geometric analogue to what the projection of a function onto a scaling function subspace is in wavelet theory. Note that each $\mathbb{P}_{j,k}(C_{j,k})$ is the best d-dimensional planar approximation to $C_{j,k}$ (in the least squares sense), which differs from the construction in [3].

Multiscale geometric wavelets. We introduce our wavelet encoding of the difference between $\mathcal{M}_j$ and $\mathcal{M}_{j+1}$, for $j \geq 0$. It is natural to define

$$x_j \equiv P_{\mathcal{M}_j}(x) := \mathbb{P}_{j,k}(x), \quad x \in C_{j,k}, \forall(j, k). \tag{2.7}$$

Again, note the difference between x_j and its equivalent in [3]. Fix $x \in C_{j+1,k'} \subset C_{j,k}$. The difference $x_{j+1} - x_j$ is a high-dimensional vector in $\mathbb{R}^D$, however it may be decomposed into a sum of vectors in certain well-chosen low-dimensional spaces, shared across multiple points, in a multiscale fashion. We proceed as follows: for $j \leq J - 1$ we let

$$\begin{aligned} Q_{\mathcal{M}_{j+1}}(x) &:= P_{\mathcal{M}_{j+1}}(x) - P_{\mathcal{M}_j}(x) \\ &= x_{j+1} - \mathbb{P}_{j,k}(x_{j+1}) + \mathbb{P}_{j,k}(x_{j+1}) - \mathbb{P}_{j,k}(x) \\ &= (I - P_{j,k})(x_{j+1} - c_{j,k}) + P_{j,k}(x_{j+1} - x). \end{aligned} \tag{2.8}$$

Define the wavelet subspace and translation as

$$W_{j+1,k'} := (I - P_{j,k})\, V_{j+1,k'}; \tag{2.9}$$

$$w_{j+1,k'} := (I - P_{j,k})(c_{j+1,k'} - c_{j,k}). \tag{2.10}$$

Clearly $\dim W_{j+1,k'} \leq \dim V_{j+1,k'} = d$. Let $\Psi_{j+1,k'}$ be an orthonormal basis for $W_{j+1,k'}$ which we call a geometric wavelet basis (see Fig. 1). Then we may rewrite (2.8) as

$$\begin{aligned} Q_{\mathcal{M}_{j+1}}(x) = \Psi_{j+1,k'}\Psi_{j+1,k'}^*(x_{j+1} - c_{j+1,k'}) + w_{j+1,k'} \\ - \Phi_{j,k}\Phi_{j,k}^*(x - x_{j+1}). \end{aligned} \tag{2.11}$$

```
GWMRA = GeometricWaveletsMRA (X, d, ε)
// Input:
// X: a set of n samples from M ⊂ R^D
// d: a dimension
// ε: a precision parameter
// Output:
// A tree T of dyadic cubes {C_{j,k}}, with local means {c_{j,k}} and
SVD bases {Φ_{j,k}}, as well as a family of geometric wavelets
{Ψ_{j,k}}, {w_{j,k}}
Construct a tree T of dyadic cubes {C_{j,k}} with centers {c_{j,k}}.
J ← finest scale with the ε-approximation property.
Let cov_{J,k} = |C_{J,k}|^{-1} Σ_{x∈C_{J,k}} (x − c_{J,k})(x − c_{J,k})^*, and com-
pute rank-d SVD(cov_{J,k}) ≈ Φ_{J,k} Σ_{J,k} Φ^*_{J,k}, for all k ∈ Γ_J.
for j = J − 1 down to 0
    for k ∈ Γ_j
        Compute cov_{j,k} and Φ_{j,k} as above
        For each k' ∈ children(j, k), construct the wavelet
        basis Ψ_{j+1,k'} and translation w_{j+1,k'}
    end
end
Set Ψ_{0,k} := Φ_{0,k} and w_{0,k} := c_{0,k} for k ∈ Γ_0.
```

Fig. 2. Pseudocode for the construction of geometric wavelets

The last term $x - x_{j+1}$ can be closely approximated by $x_J - x_{j+1} = \sum_{l=j+1}^{J-1} Q_{\mathcal{M}_{l+1}}(x)$ as the finest scale $J \to +\infty$, under general conditions. This equation splits the difference $x_{j+1} - x_j$ into a component in $W_{j+1,k'}$, a translation term that only depends on (j,k) and lies in the orthogonal complement of $V_{j,k}$ in $\mathbb{R}^D$ but not necessarily in $W_{j+1,k'}$, and a projection onto $V_{j,k}$ of a sum of differences $x_{l+1} - x_l$ at finer scales.

The two-scale relationship, by definition of $Q_{\mathcal{M}_{j+1}}$,

$$P_{\mathcal{M}_{j+1}}(x) = P_{\mathcal{M}_j}(x) + Q_{\mathcal{M}_{j+1}}(x), \tag{2.12}$$

may be iterated across scales:

$$x = P_{\mathcal{M}_j}(x) + \sum_{l=j}^{J-1} Q_{\mathcal{M}_{l+1}}(x) + (x - P_{\mathcal{M}_J}(x)) \,. \tag{2.13}$$

The above equations allow to efficiently decompose each step along low-dimensional subspaces, leading to an efficient encoding algorithm (see Fig. 2). We have therefore constructed a multiscale family of projection operators $P_{\mathcal{M}_j}$ (one for each node $C_{j,k}$) onto approximate local tangent planes and detail projection operators $Q_{\mathcal{M}_{j+1}}$ (one for each edge) encoding the differences, collectively referred to as a GWMRA structure. The cost of encoding the GWMRA structure is dominated by that of the scaling functions $\{\Phi_{j,k}\}$, which is $O(dD\epsilon^{-\frac{d}{2}})$, and the time complexity of the algorithm is $O(Dn\log(n))$ [2].

Theorem 2.1 (Geometric Wavelet Decomposition)**.** *Let $(\mathcal{M}, g)$ be a $\mathcal{C}^2$ manifold of dimension d in $\mathbb{R}^D$. Let $\{P_{\mathcal{M}_j}, Q_{\mathcal{M}_{j+1}}\}$ be a GWMRA. For any $x \in \mathcal{M}$, there exists a constant $C = C(x)$ and a scale $j_0 = j_0(\mathrm{reach}(\mathbb{B}_1(x)))$, such that for any $j \geq j_0$,*

$$\|x - P_{\mathcal{M}_{j_0}}(x) - \sum_{l=j_0+1}^{j} Q_{\mathcal{M}_l}(x)\| \leq C \cdot 2^{-2j}. \tag{2.14}$$

```
{q_{j,x}} = FGWT(GWMRA, x)
// Input: GWMRA structure, x ∈ M
// Output: A sequence {q_{j,x}} of wavelet coefficients
for j = J down to 0
    q_{j,x} = (Ψ^*_{j,x} Φ_{j,x}) Φ^*_{j,x}(x − c_{j,x})
end
```

```
x = IGWT(GWMRA, {q_{j,x}})
// Input: GWMRA structure, wavelet coefficients {q_{j,x}}
// Output: Reconstruction x at scale J
Q_{M_J}(x) = Ψ_{J,x} q_{J,x} + w_{J,x}
for j = J − 1 down to 1
    Q_{M_j}(x) = Ψ_{j,x} q_{j,x} + w_{j,x} − P_{j−1,x} Σ_{ℓ>j} Q_{M_ℓ}(x)
end
x = Ψ_{0,x} q_{0,x} + w_{0,x} + Σ_{j>0} Q_{M_j}(x)
```

Fig. 3. Pseudocodes for the Forward and Inverse GWTs

Geometric Wavelet Transforms (GWT). Given a GWMRA structure, we may compute a discrete Forward GWT for a point $x \in \mathcal{M}$ that maps it to a sequence of wavelet coefficient vectors:

$$q_x = (q_{J,x}, q_{J-1,x}, \ldots, q_{1,x}, q_{0,x}) \in \mathbb{R}^{d+\sum_{j=1}^{J} d^w_{j,x}} \tag{2.15}$$

where $q_{j,x} := \Psi^*_{j,x}(x_j - c_{j,x})$, and $d^w_{j,x} := \mathrm{rank}(\Psi_{j,x}) \leq d$. Note that, for a fixed precision $\epsilon > 0$, q_x has a maximum possible length $(1 + \frac{1}{2}\log_2 \frac{1}{\epsilon})d$, which is independent of D and nearly optimal in d [3]. On the other hand, we may easily reconstruct the point x at all scales using the GWMRA structure and the wavelet coefficients, by a discrete Inverse GWT (Fig. 3 displays the pseudocodes for both transforms).

3. A TOY DATASET

We consider a toy example of a 2-dimensional *swissroll* manifold in $\mathbb{R}^{100}$ and apply the algorithm to the sampled data (6000 points, without noise) in Fig. 4. The resulting wavelet coefficients matrix is very sparse (with about 40% of the coefficients below 1 percent of the maximum magnitude). The reconstructed manifolds also approximate the original manifold well, with the errors decreasing quadratically with respect to scale.

4. VARIATIONS AND OPTIMIZATIONS

We presented a plain construction of the geometric wavelets for the data sampled from a manifold of known dimension d. We mention several variations and optimizations, aimed at reducing the cost of encoding the geometric wavelet dictionary and/or speeding up the decay of the wavelet coefficients [2, 3].

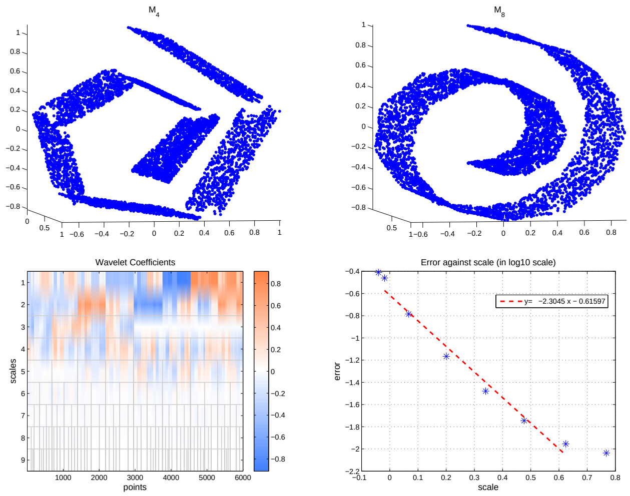

Fig. 4. Top: Reconstructions of a *swissroll* at scales 4 (left) and 8 (right), using 6000 random samples, by the geometric wavelets. Bottom left: matrix of wavelet coefficients. The x-axis indexes the points (arranged according to the tree), and the y axis indexes the scales from coarse (top) to fine (bottom). Note that each block corresponds to a different subset of wavelet bases. Bottom right: reconstruction errors as a function of scale (both in $\log_{10}$ scale).

Locally adaptive dimensions [2]. Currently we use the same dimension d everywhere in the tree when constructing geometric scaling functions for data sampled from a d-dimensional manifold. We may use local dimensions $d_{j,k}$ that are adapted to the cubes $C_{j,k}$, extending the construction to any point cloud (see Fig. 5).

Orthogonal geometric wavelets [2]. Neither the vectors $Q_{\mathcal{M}_{j+1}}(x)$, nor any of the terms that comprise them, are in general orthogonal across scales. On one hand, this is natural since $\mathcal{M}$ is nonlinear; on the other hand, the $Q_{\mathcal{M}_{j+1}}(x)$ may be almost parallel across scales or, the subspaces $W_{j+1,k'}$ may share directions across scales. We may then more efficiently encode the dictionary by not encoding twice shared directions, leading to orthogonal geometric wavelets.

Splitting of the wavelet subspace $W_{j+1,k'}$ [3]. We may split $W_{j+1,k'}$ into a part that depends only on (j,k) and another on $(j+1,k')$ by defining $W_{j,k}^{\cap} := \cap_{k' \in \mathrm{children}(j,k)} W_{j+1,k'}$ and letting $W_{j+1,k'}^{\perp}$ be the orthogonal complement of $W_{j,k}^{\cap}$ in $W_{j+1,k'}$, so that we only need to encode $W_{j,k}^{\cap}$ once for all children. This is particularly efficient for encoding the wavelet dictionary when $\dim(W_{j,k}^{\cap})$ is large relative to all $\dim(W_{j+1,k'})$.

A fine-to-coarse strategy without tangential corrections [3]. In this variation, we can remove the tangential term in (2.11) by using the sequence $\tilde{x}_j = \mathbb{P}_{j,x}(\tilde{x}_{j+1})$, for $j < J$, and $\tilde{x}_J := x_J$, instead of the previous one $x_j = \mathbb{P}_{j,x}(x)$, so that we do not need to encode the scaling functions $\{\Phi_{j,k}\}$ (see [3] for details).

Pruning of the geometric wavelet tree [2]. Even with the above techniques, our construction of the tree is still not optimal. One may consider pruning the tree to minimize the overall encoding cost. See Fig. 6 for an illustration.

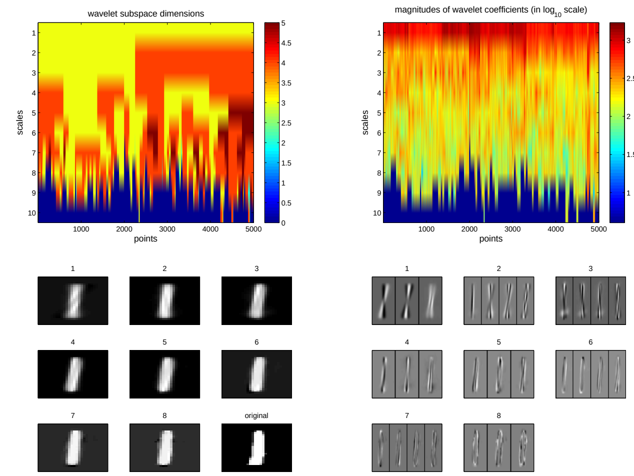

Fig. 5. We sampled 5000 handwritten digits 1 from the MNIST database [6], and used locally adaptive dimensions by keeping 50% variance in each cube. Top: dimensions of all wavelet subspaces (left) and magnitudes (in $\log_{10}$ scale) of the wavelet coefficients (right); Bottom: Approximations of a digit 1 (left) and elements of the wavelet dictionary used in the expansions (right).

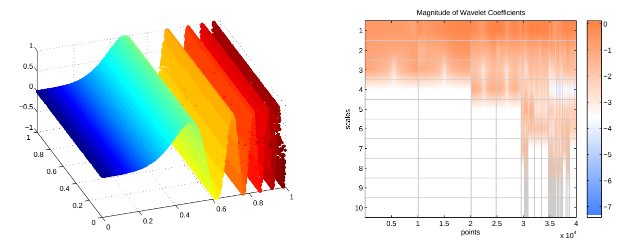

Fig. 6. Left: a manifold of nonuniform smoothness, sampled at 40,000 points. Right: compressed wavelets tree and coefficients.

5. REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: Design of dictionaries for sparse representation. In *Proc. of SPARS 05'*.

[2] W. K. Allard, G. Chen, and M. Maggioni. Multiscale geometric methods for data sets II: Geometric wavelets. *Submitted, 2011*.

[3] G. Chen and M. Maggioni. Multiscale geometric wavelets for the analysis of point clouds. In *Proc. of the 44th Conference on Information Sciences and Systems (CISS)*, pages 1-6, 2010.

[4] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *PNAS*, 102(21):7426-7431, 2005.

[5] G. David. *Wavelets and Singular Integrals on Curves and Surfaces*. Springer-Verlag, 1991.

[6] Y. LeCun and C. Cortes. The MNIST database of handwritten digits. Available online at *http://yann.lecun.com/exdb/mnist/*.

[7] A. V. Little, M. Maggioni, and L. Rosasco. Multiscale geometric methods for data sets I: Intrinsic dimension. *Submitted, 2011*.

[8] A. Szlam and G. Sapiro. Discriminative k-metrics. In *ICML 09'*.