Perspectives and Challenges to Harmonic Analysis and Geometry in High Dimensions: Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data

R.R. Coifman*

Department of Mathematics, Yale University, New Haven, CT 06520, USA
coifman@math.yale.edu

1 Introduction

Our goal is to describe some of the mathematical challenges confronted, when dealing with massive data sets. We point out that the trend set in hard analysis by L Carleson, of integrating combinatorics with geometry and harmonic analysis, is a powerful guide in the context of analytic geometry of data.

The task confronted by the scientist or engineer is to organize and structure enormously complex high dimensional clouds of points. More importantly, the scientist is challenged by the need to approximate empirical functions depending on many parameters effectively . Here effective methods are the ones for which the *complexity of the method does not grow exponentially with the dimension. In effect no tools from classical approximation theory and analysis exist.*

Various methods for embedding high dimensional data sets in relatively modest dimensions were introduced by various groups of researchers. In particular Johnson and Lindenstrauss have shown that N points in N dimensions can almost always be projected to a space of dimension $C \log N$ with control on the ratio of distances. J Bourgain has proved that any finite metric space with N points can be embedded by a bi-Lipschitz map into $\log N$ dimensional Euclidean space with bi-Lipschitz constant $\log N$. Various randomized versions of this theorem have become useful tools for protein mapping and other data analysis tasks. These results indicate that in practice to address the problem of functional approximation it would be enough to restrict our attention to dimensions below 50.

* The results in this talk are joint work with S. Lafon, A. Lee, M. Maggioni, S. Zucker.

2 Diffusion Geometries and Data Organization

The situation is not so difficult when dealing with clouds of points in high dimensions which are distributed near lower dimensional manifolds or varifolds. In particular empirical manifold learning is an active area of research.

We now would like to use Harmonic analysis as a tool to process empirical data sets.

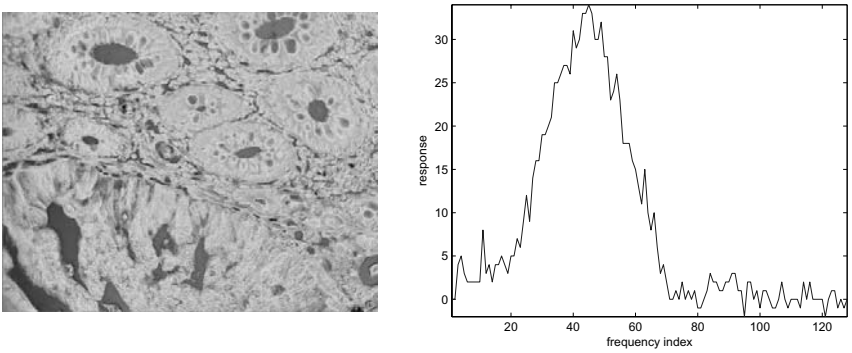

Fig. 1. Hyperspectral image of a pathology slide (left) with associated spectrum (right).

The hyperspectral image of a pathology slide to the left in Fig. 1 provides an illustration of some of the issues mentioned above, here to each pixel of the image is associated a spectrum (Fig. 1, right) representing the electromagnetic absorption of the tissue in 28 bands. This spectrum reflects the mix of biological constituents. The coloring in RGB of the image provides at each location a mix of three tissue types.

For physical reasons it is expected that these points in 28 dimensional space lie on different low dimensional submanifolds.

Our goal is to provide a general methodology to describe and parametrize it. Our method provides a framework for structural multiscale harmonic analysis on subsets (data) of $\mathbb{R}^n$ and on graphs. We use diffusion semigroups to generate multiscale geometries in order to organize and represent complex structures. We build the diffusions through an "infinitesimal" Markov process, and will show that the top eigenfunctions of the Markov operator permit a low dimensional geometric embedding of the data set into $\mathbb{R}^n$ so that the ordinary Euclidean distance in the embedding space measures intrinsic diffusion metrics on the data. Moreover we will indicate how empirical functions on the data can be naturally extended to all of space.

While some of these ideas appear in a variety of contexts of data analysis, such as spectral graph theory, manifold learning, nonlinear principal compo-

nents and kernel methods. We augment these approaches by showing that the diffusion distances are key intrinsic geometric quantities linking spectral theory of the Markov process (Laplace operator, or Kernels) to the corresponding geometry of the data, relating localization in spectrum to localization in space, opening the door to the application of methods from signal processing to analyze functions and transformations on the data.

Initially, our goal is to describe efficiently (empirical) functions on a set Γ (data) or on a discrete graph. In particular we consider the analysis of restrictions of band limited functions to the data (i.e. functions whose Fourier transform is supported in a ball B). Specifically, the space of restrictions of band limited function to the data set Γ, is spanned by the eigenvectors of the covariance operator

$$k_B(x,y) = \int_B e_\xi(x) e_\xi(-y) \mathrm{d}\xi \ ,$$
$$\lambda_j \varphi_j = \int_\Gamma k_B(x,y) \varphi_j(y) \mathrm{d}y \ .$$

These eigenfunctions have natural extensions as band limited functions in $\mathbb{R}^n$ given by

$$\phi_j(x) = \frac{1}{\lambda_j} \int_\Gamma k_B(x,y) \varphi_j(y) \mathrm{d}y \ .$$

It turns out that $\phi_j(x)$ form an orthogonal set on $\mathbb{R}^n$ of band limited functions whose norm is maximized on the set (generalizing the classical prolate functions). For the example of the unit sphere these eigenfunctions are the spherical harmonics which are the eigenfunctions of the Laplace operator. We will refer to the extensions of these eigenfunctions as geometric harmonics (since these extensions are characterized as the minimal norm band limited extension of the given restriction to the set). We will use these extensions for estimating empirical regression off the data set.

More generally for data lying on a submanifold of $\mathbb{R}^n$, any restriction of a positive radial kernel leads to approximations of eigenfunctions of the Laplace–Beltrami operator on the manifold. We now extend results of Belkin et al. relating kernels and Laplace–Beltrami operators on submanifolds of Euclidean space.

We restrict a positive symmetric kernel to the data set, as an operator (or matrix)

$$K(f) = \int_\Gamma k(x,y) f(y) \mathrm{d}y$$

and diagonalize K as

$$k(x,y) = \sum \lambda_i^2 \varphi_i(x) \varphi_i(y) \ .$$

Then

$$D^2(x,y) = k(x,x) + k(y,y) - 2k(x,y) = \sum \lambda_i^2 \big(\varphi_i(x) - \varphi_i(y)\big)^2$$

is the square of the metric $D(x,y)$.

If the kernel is given as a function of some initially given metric $d(x,y)$ (for example d could be the geodesic metric), and $k(x,y) = k(d^2(x,y))$, $k'(0) = 1$, then

$$D(x,y) \approx \frac{d(x,y)}{1+d(x,y)} = d_0(x,y) \ .$$

This observation shows that for data on a compact submanifold of $\mathbb{R}^n$, the kernel metric is equivalent to the original metric.

A particularly important case arises from a scaled weighted Gaussian kernel operator (by a density of point distribution $p(y)$)

$$\int_\Gamma \exp(-d(x,y)^2/\varepsilon) f(y) p(y) \mathrm{d}y \ .$$

This kernel has to be renormalized as follows; let

$$p_\varepsilon(x) = \int_\Gamma \exp(-d(x,y)^2/\varepsilon) p(y) \mathrm{d}y$$

and

$$\nu_\varepsilon(x) = \int_\Gamma \exp(-d(x,y)^2/\varepsilon) \frac{p(y)\mathrm{d}y}{p_\varepsilon(x)} \ .$$

Then the operator[2]

$$A_\varepsilon(f) = \int_\Gamma a_\varepsilon(x,y) f(y) p(y) \mathrm{d}y \ ,$$

where

$$a_\varepsilon(x,y) = \frac{\exp(-d(x,y)^2/\varepsilon)}{\nu_\varepsilon(x) p_\varepsilon(y)}$$

is an approximation to the Laplace–Beltrami diffusion kernel at time ε.

The operator A_ε can be used to define a discrete approximate Laplace operator

$$\Delta_\varepsilon = \frac{1}{\varepsilon}(A_\delta - I) = \Delta_0 + \sqrt{\varepsilon} R_\delta \ ,$$

where R is bounded on band limited functions and Δ_0 is the Laplace–Beltrami operator on the manifold. From this we can deduce the following theorem:

Theorem 2.1.

$$(A_\delta)^{t/\delta} = (I + \varepsilon\Delta_\delta)^{t/\delta} = (I + \varepsilon\Delta_0)^{t/\delta} + O(\sqrt{\varepsilon}) = \exp(t\Delta_0) + O(\sqrt{\varepsilon})$$

and the kernel of $(A_\delta)^{t/\delta}$ *is given as*

$$a_t(x,y) = \sum \lambda_i^{2t/\delta} \psi_i(x) \psi_i(y) = \sum \exp(-\mu_l t) \phi_l(x) \phi_l(y) + O(\sqrt{\varepsilon}) \ ,$$

where the ϕ *are the eigenfunctions of the limiting Laplace operator and all estimates are relative to any fixed space of band limited functions.*

[2] Shoenberg proved that this operator is positive, if and only if the metric d embeds isometrically in Hilbert space.

Strictly speaking we assume here that the data is relatively densely sampled (each ball of radius ε contains several points) on a closed compact manifold. In case the data only covers a sub domain of the manifold, the Laplace operator needs to be interpreted as the restriction of the Laplace operator with Neumann boundary condition.

The fundamental observation is that the numerical rank of the powers of A decreases rapidly (see Fig. 2 below) and therefore the diffusion distance given by

$$a_t(x,x) + a_t(y,y) - 2a_t(x,y) = D_t^2(x,y)$$

can be computed to high accuracy using only the corresponding eigenfunctions. This choice of an embedding into Euclidean space so as to convert diffusion distance on the manifold into Euclidean distance in the embedding will be called a diffusion map.

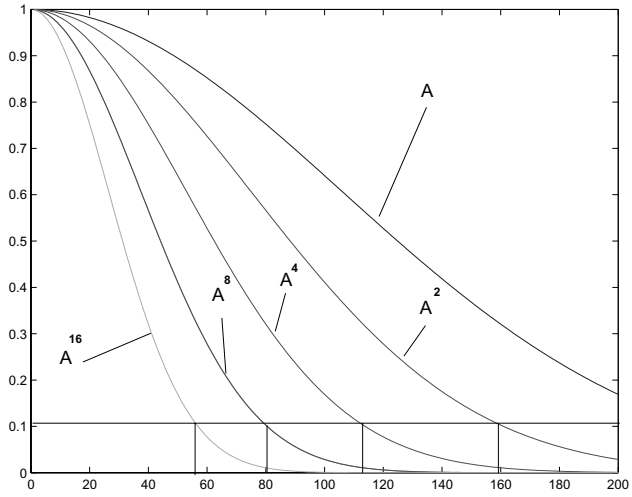

Fig. 2. Some examples of the spectra of powers of A.

We illustrate this point for the case of a closed rectifiable curve, for which the first two non constant eigenfunctions give a realization of the arc length parametrization onto the circle of the same length. In fact a simple computation relative to the arc length parametrization shows that the heat kernel is given by

$$e_t(x,y) = \sum \exp(-k^2 t) \cos\big(k(x-y)\big)$$

and

$$\mathrm{e}^t D_t^2(x,y) = |\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{\mathrm{i}y}|^2 \big(1 + \mathrm{e}^{-3t} r_t(x,y)\big) \ ,$$

where r is bounded.

In Fig. 3 we see points distributed (non uniformly) on the spiral on the left, the next embedding into the plane is given by the conventional graph Laplacian normalization, while the circle was obtained as above. The graph on the right is the density of points on the circle.

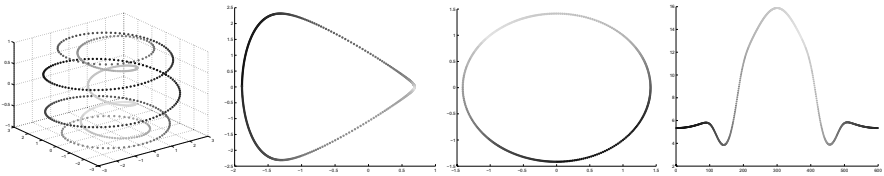

Fig. 3. From left to right: points distributed on a spiral—an embedding into the plane—a circle obtained as above—a graph depicting the density of points on the circle.

More generally we have the following theorem

Theorem 2.2. *Let $x_i \in \Gamma \subseteq M$ be a data set in the compact Riemannian manifold M so that each point in the manifold is at a distance ε from one of the data points and let the matrix $a_t(x,y)$ be defined as above on the data, with*

$$a_t(x,y) = \sum \lambda_i^{2t/\varepsilon} \psi_i(x)\psi_i(y) \ .$$

Then there exists an m such for all t sufficiently large the diffusion map is an embedding of M into m dimensional Euclidean space which is approximately isometric relative to the extrinsic Euclidean distance.

$$x \in M \mapsto \tilde{x} = \big(\lambda_1^{t/\varepsilon}\psi_1(x), \lambda_2^{t/\varepsilon}\psi_2(x), \dots \lambda_m^{t/\varepsilon}\psi_m(x)\big) \in \mathbb{R}^m$$

$$D_t^2(x,y) = \sum_{i=1}^{m} \lambda_i^{2t/\varepsilon}\big(\psi_i(x) - \psi_i(y)\big)^2\big(1 + O(\mathrm{e}^{-\alpha t})\big) = |\tilde{x} - \tilde{y}|^2\big(1 + O(\mathrm{e}^{-\alpha t})\big)$$

The proof of this theorem uses the fact that for small t and large m we have an embedding, we then pick the smallest m for which we have a bi-Lipschitz embedding and the next eigenvalue is strictly smaller (we can also maximize the spectral gap to have higher precision).

The next example (Fig. 4) embeds an hourglass surface into three dimensional Euclidean space so that the diffusion distance in embedding space between two points is the length of the chord connecting them.

Since the diffusion is slower through the bottle neck the two components are farther apart in diffusion metric.

Figure 5 illustrates the organizational ability of the diffusion maps on a collection of images that was given in random order as reordered by the mapping given by the first two nontrivial eigenfunctions.

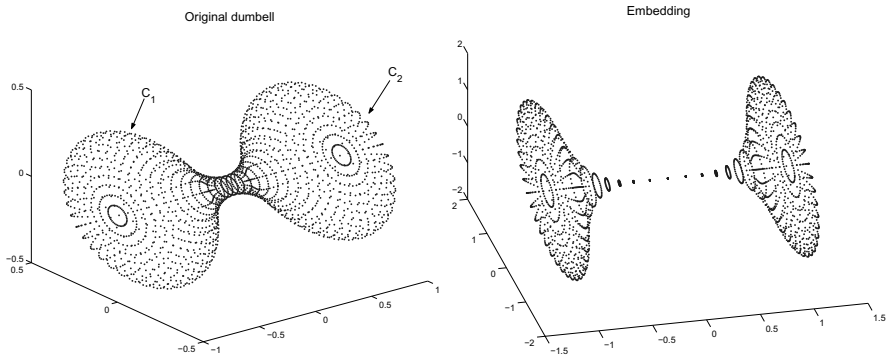

Fig. 4. Original dumbbell (left) and embedding (right).

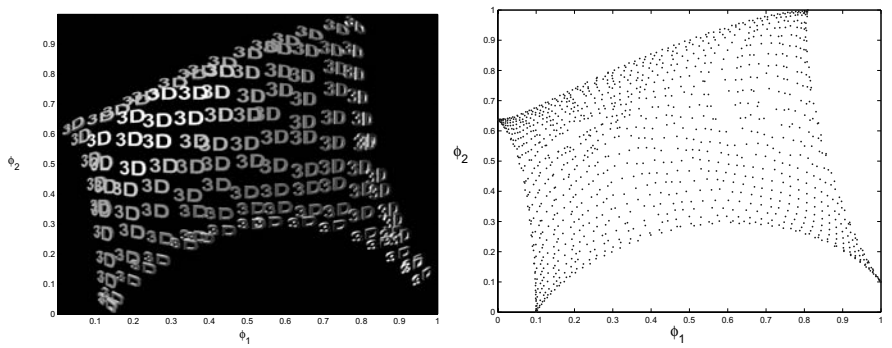

Fig. 5. The first two eigenfunctions organize the small images which were provided in random order.

Figure 6 shows the conventional nearest neighbor search compared with a diffusion search. The data is a pathology slide, each pixel is a digital document (spectrum below for each class).

3 Extension of Empirical Functions off the Data Set

An important point of this multiscale analysis involves the relation of the spectral theory on the set to the localization on and off the set of the corresponding eigenfunctions. In the case of a smooth compact submanifold of Euclidean space it can be shown that any band limited function of band B can be expanded to exponential accuracy in terms of eigenfunctions of the Laplace operator with eigenvalues μ^2 not exceeding CB.

Conversely every eigenfunction of the Laplace operator satisfying this condition extends as a band limited function with band $C'B$ (both of these state-

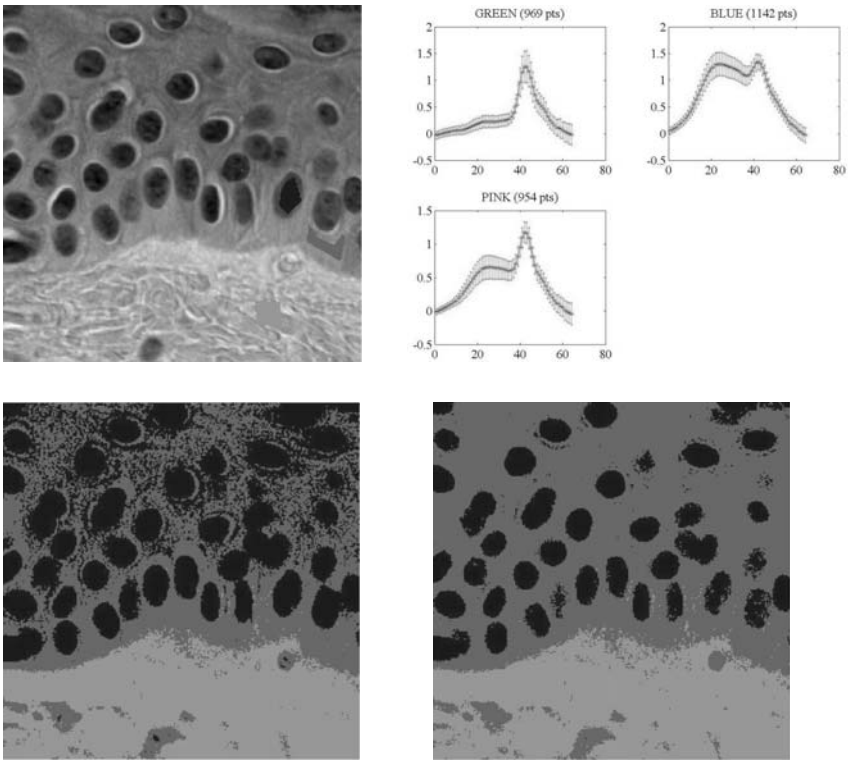

Fig. 6. Top-left to bottom-right: 256×256 image with 861 labeled points—spectrum for each class–nearest neighbor search—extension to all points.

ments can be proved by observing that we can estimate the size of a derivatives of order $2m$ of eigenfunctions of the Laplace operator as a power m of the eigenvalue). If we extend the eigenfunctions as constant in the normal direction to the submanifold (by introducing an appropriate smooth partition of unity in a neighborhood of the submanifold). It is easy to see that the Fourier transform at ξ of such an extension is estimated by $C_m(\mu/|\xi|)^m$ and this shows that eigenfunctions of the Laplace operator corresponding to eigenvalue μ on the manifold are well approximated by restrictions of band limited functions of band C_μ.

We conclude that given an empirical function on the manifold which can be approximated to some accuracy with eigenfunctions whose frequencies are localized, or which is expanded in terms of a multiscale basis involving eigenvalues of the Laplace operator not exceeding μ, then such a function can be extended as a band limited function off the set to a distance corresponding

to $C\mu^{-1}$ and the best band limited approximation can be obtained from the corresponding Band limited projection kernel $k_B(x, y)$ defined above.

References

1. M. Belkin, P. Niyogi: Neural Computation **15**, 1373 (2003)
2. F. Chung: *Spectral graph theory* (CBMS-AMS 1997)
3. R.R. Coifman, S. Lafon: Diffusion maps, submitted to Applied and Computational Harmonic Analysis (2004)
4. M. Szummer, T. Jaakkola: Partially labeled classification with Markov random walks. In: *Advances in Neural Information Processing Systems 14*, ed by T. Diettrich, S. Becker, Z. Ghahramani (University Press, United States 2001) p 945