Research Statement

Arthur Szlam

Introduction

It has become well known that harmonic analysis on graphs can be a powerful tool for attacking a wide variety of real world problems, such as image segmentation [SM00] and semi-supervised learning [Bel03]. Although the data from the problem may be presented in a high dimensional ambient space, harmonic analysis allows an intrinsic study of the data that makes use of the fact that it often has a much lower dimensional internal structure. Much work has focused on applying the analogues of the most classical Fourier analysis, using eigenfunctions of the Laplacian on the graph as the major tool.

Recently (e.g. [MC]), there have been attempts to generalize and apply the multiscale and local methods of harmonic analysis, which have been so successful in traditional signal processing problems, to the graph setting. In my work, I have continued in this direction, studying the applications of diffusions on graphs to signal processing and machine learning problems, and constructing analogues of the local cosine basis on graphs. I believe there are a powerful set of tools, some of which are just starting to be utilized, and some of which are almost available, which will impact problems accessible to harmonic analysis on graphs the same way that wavelets impacted signal processing. I hope to contribute to the discovery and application of these tools.

Discussion of Thesis work

Fourier analysis on a graph G with symmetric weight matrix W refers to the study of the eigenvectors and eigenvalues of the normalized Laplacian on the graph, as in [Chu97]. The graphs that I am interested in often arise from a dataset in a high dimensional Euclidean space, where the structure of the dataset is assumed to depend on fewer free parameters than the ambient space. For example, if one is interested in finding an algorithm to recognize handwritten digits, the dataset would be a large collection of scanned images. Suppose each scanned image was 30 pixels by 30 pixels. Then each data point would lie in 900 dimensional Euclidean space, even though we should expect that the number of parameters that the various digits depend on is far fewer than 900.

Given a dataset of points x_j in $\mathbb{R}^n$, a popular construction of a weighted graph with weight matrix W is to let

$$W(j,k) = e^{-\frac{d(x_j,x_k)^2}{t}}.$$

Let

$$D(j,j) = \sum_k W(j,k).$$

Depending on whether you are multiplying from the left or right, one thinks of the matrix $D^{-1}W$ as an (isotropic) diffusion on the dataset which has run for time t, or averaging heat kernel at time t.

Anisotropic Diffusions

One of the simplest methods of denoising signals in $\mathbb{R}^n$ is to low pass filter their Fourier transform, or equivalently, simply smooth them with an averaging convolution kernel. The same method can be used to denoise functions on a graph, using the normalized weight matrix. In either the Euclidean or the more general setting, this is not a good method because often detail of interest is lost with the noise.

In image processing, many researchers have suggested smoothing with an anisotropic diffusion which depends on the noisy image I_0, which smooths more in directions parallel to the level lines of the image, so that less image structure is lost in the filtering [ROF92, Tsc02] . The anisotropic diffusion is obtained by choosing a smoothness functional F (e.g. $F(I) = \int |\nabla I|$), and considering gradient descent on the problem of minimizing the value of F subject to constraints forcing the minimum to be in some sense close to I_0 (e.g. $\int |I - I_0|^2 < \sigma^2$).

Part of my work as a graduate student has been to show how the same basic philosophy can be used in more general situations. To illustrate, consider again the image denoising problem. Now, however, instead of finding a diffusion by stepping towards the minimum of a smoothness functional on the space of images, we can introduce a useful anisotropy by adding some number of "feature coordinates" (which depend on the image to be denoised) to the cartesian coordinates of the image, and finding the isotropic diffusion on the graph with edge weights constructed in the image's "feature space". A standard way to build the feature space is to consider the image as a function $I : \mathbb{R}^2 \mapsto \mathbb{R}$ and choose d filters (i.e. functions $f_1 ... f_d$ on $\mathbb{R}^2$), and so get a mapping $\mathbb{R}^2 \mapsto \mathbb{R}^d$ defined by $x \mapsto (I * f_1(x), ..., I * f_d(x))$. There are many possibilities for the choice of filters, for example one could take a few wavelet functions at different scales, or edge filters. The same construction is available on a sampled image, using a discrete convolution. Once the graph and diffusion have been built, the image can be considered as a function on this graph, and smoothed by the diffusion to denoise it.

Figure 1: A noisy and denoised image. 49 7×7 filters were used, each having one entry equal to 1, and the rest 0, so the resulting graph consists of all 7×7 image patches.

The power of the technique I have described is that it does not require the underlying dataset to be Euclidean, or even have any differential structure. For example, we can use the idea of adding anisotropy with feature coordinates in statistical learning problems. Several authors have discussed semi-supervised learning in the context of diffusions on a graph, see [Bel03] . The semi-supervised learning problem is as follows: one is given a large number of datapoints, and some of these datapoints are labeled. The aim is to label the remaining points. For example, the dataset could be the scanned digits from above, and the classes would be defined by what digit a particular image is. The problem would then be to guess which digit each image is given a few correctly labeled images. Under the assumption that datapoints which are very close together are similar (but not necessarily the converse), harmonic analysis has been shown to be an effective tool in attacking these problems. For each class c, one considers the function $\chi_c^{labeled}$ which is 1 on the labeled points of class c, and 0 elsewhere. In [Bel03], the eigenfunctions of the Laplacian on a graph defined by the dataset are used to filter the $\chi_c^{labeled}$, obtaining functions f_c. Unlabeled points x are then classified by $arg \max_c f_c(x)$, and the f_c can be considered confidence measures for the classifiers. I have found that using an anisotropic diffusion as above consistently and significantly improves the classifiers obtained. For feature coordinates, one can simply use the confidence functions from the original classifier. The new diffusion then mixes faster along the boundaries of the suspected classes, but mixes slower across the suspected boundaries. In some sense, here we are denoising a noisy class function.

Local Bases

In Euclidean spaces, multiscale bases are a powerful tool for signal analysis. Recently, building on ideas from Stein [Ste70], Maggioni and Coifman have developed a generalization of the wavelet construction to graphs using powers of the diffusion on the graph as a replacement for scaling in Euclidean space [MC]. In their construction, the frequencies of the operator are broken into dyadic blocks, and then bases of each frequency block are constructed with an aim towards minimizing the support of the basis elements. In [MCBS05], the the construction of wavelet packets in this framework was explored.

I have used the dual approach, where the space side is decomposed by hand, to construct local bases on graphs, i.e. the analogue of the local cosine construction. An important part of the construction is then of course the decomposition of the graph into nice pieces, which is a serious problem by itself. There are several possible methods for this; a standard method is dividing the graph into the nodal sets of the first (or the first few) nontrivial eigenfunction of the Laplacian on the graph (spectral clustering). Once such a decomposition method has been fixed, it can be iterated to give a multiscale partition of the graph.

Given a partition of a space with some sort of smoothness structure (say a manifold), one can ask how to build smooth bases whose elements are supported mostly on a single element in the partition. In one dimension, the local cosine construction uses a rising cutoff function and the reflection $x \mapsto -x$ to provide an answer [CM91]. Kovacevic and Bernardini generalized this construction to a space and partition of the space where a group of smooth symmetries corresponds points in adjacent sets of the partition [BK00, BB99]. However, even for smooth manifolds, for many reasonable partitions, such a group is not readily available. One does often have a reasonable definition of a distance to the boundary of a set in the partition, which can be used to build a reflection across the boundary. By phrasing the correct eigenproblem, I showed how one can build an analogue of the local cosine bases using sequential binary partitions of the space and some sort of

Laplacian or heat operator. This construction experimentally is robust enough to work on discrete sets which in some way approximate a smooth space, where there is only a "rough" reflection or even just rough distances to the boundaries between siblings in the multiscale partition [SMBC05].

The construction is not yet practical because of the lack of fast algorithms. However, in applications, I have demonstrated that one can still profitably use unsmoothed local bases, which are analogous to the block DCT used in JPEG, or nonorthogonal frames of smoothed block cosines or smoothed Haar functions [SMBC05].

Future Work

My general objective will be to continue to try to generalize and apply the tools of harmonic analysis to problems amenable to study in the framework of graphs arising from low dimensional data in high dimensional ambient space. More specifically, I would like to consider the following:

• There are several problems that seem approachable using local bases built on graphs. I would like to try to develop an image compression algorithm using a patch graph as described above, but built on a very compressed version of the original image, for example, just storing the edges or very few large wavelet coefficients. Partitioning filtered image graphs is a good approach to image segmentation, e.g. [SM00]; therefore, the local cosines built from the patch graph follow the contours of the (compressed) image. By judiciously choosing the initial compression and the parameters of the local bases, I believe it will be possible to compress images very efficiently.

I think the local bases will also be useful in studying functions on state spaces of dynamical systems. For example, the system could be a molecule, the state space could be the possible configurations of the molecule, and the functions on the molecule could be strain energies or likelihood of bonding to a different molecule in various configurations. At present, there seems to be very few good tools available for multiscale analysis with local basis functions, and for very large datasets, the block cosines or block haar functions seem to the only multiscale analyses computationally possible.

• A fast algorithm for the smooth local bases is desirable. While I do not currently see any way of finding a fast transform (in the Euclidean case, one uses the FFT for the fast local cosine transform), I think it should be possible to at least build the smooth multiscale projections fast, and to do so without needing any reflections, just distances to boundaries.

• I would like to use the anisotropic graph diffusion for video denoising. A fast and strong video denoising algorithm would have several immediate applications, for example making it possible to do real time MRI. The problem here is finding the correct way to treat the time variable.

• In the construction of the graph weights from datapoints there is very little theory on how to choose the function of the distances. The most popular choice, $W(j,k) = e^{-\frac{d(x_j,x_k)^2}{t}}$, already has the global parameter t. The choice of t is extremely important but still poorly understood. One problem is that there is no reason that a given dataset should be homogenously distributed; and so there simply may be no good choice of t. There have been some attempts to try to choose different t at different locations; i.e. use some sort of local statistics to allow the diffusion to run faster or slower on different parts of the graph, for example, see [ZMP04], where the distance between two points was measured in the scale of the geometric mean of kth nearest neighbor from each of the points. There are other, more subtle ways to measure the local scale around a point (for example, local PCA), and so there seems be a lot of room for improvement in this matter.

• Like the sine and cosine functions on the interval, the eigenfunctions of the Laplacian on a compact manifold are supported on the whole manifold. However, experimentally, on sampled manifolds, the eigenfunctions of Laplacian tend to have their energy well localized, especially in the high frequencies. A special case of this effect can be seen in a “noisy” convolution with a function with small support, i.e. a an operator which has a banded matrix form, where the band consists of a translated function plus noise. With sufficient noise, these operators (which are like convolutions on a noisy circle) have eigenfunctions with energy localized in a small segment. It would be very interesting to understand why this happens, and how we can make use of this effect. One application could be the design of wavelets on graphs by choosing an appropriate kernel, and diagonalizing it.

• In the construction of local bases, one first has to break up the graph into reasonable pieces. As mentioned, one method is to repeatedly subdivide the graph into the nodal sets of the first non-trivial eigenfunction (spectral clustering). There are some results describing the regularity of the partitions obtained for general graphs, eg [KVV04]. On the other hand, it seems very little has been proved for the asymptotic behavior of the partitions (as the number of subdivisions increases) on smooth sets. In fact, if we just consider domains in $\mathbb{R}^2$, almost nothing is known about the asymptotic behavior of the spectral clustering of the domain, and surprisingly, until very recently, little was known about the geometry of even one split by the first eigenfunction [BB99]. In [SMBC05], we make several conjectures about the asymptotic regularity of the partition. The most extreme is that if we always take the partition according to the Neumann nodal domains with shortest boundary (in the case that the eigenvalue of the domain is not simple), the partition will limit in an appropriate sense to either squares, isosceles right triangles, or polar sectors. A milder conjecture, which seems reachable, is that there is a constant C_D depending on the initial domain so that the chord-arc constant of any set in the partition is smaller than C_D (the chord arc constant of a domain D is $min_{x,y \in \partial D} l(x,y)/d(x,y)$, where $l(x,y)$ is the length along the boundary. In computer experiments the spectral decomposition of a domain satisfies the strong conjecture.

Conclusion

The techniques which I have studied and would like to continue to improve on are useful for organizing and analyzing data which is noisy and high dimensional, and has a complicated global structure. The only assumption necessary to apply these techniques is that locally some lower dimensional structure exists; the algorithms automatically find that structure and make use of it. There are a huge number of potential applications, including image/video compression and denoising, statistical learning, and the analysis of dynamical systems. Often, in addition to being strong and fast, the algorithms which result from these methods are particularly simple, and easy to implement. There is currently a great research opportunity for people who have a good understanding of modern harmonic analysis and are interested in applications. I hope to make the the most of this opportunity in the next few years.

References

[BB99] Rodrigo Bañuelos and Krzysztof Burdzy. On the "hot spots" conjecture of J. Rauch. *J. Funct. Anal.*, 164(1):1–33, 1999.

[Bel03] M. Belkin. *Problems of learning on manifolds.* PhD thesis, University of Chicago, 2003.

[BK00] Riccardo Bernardini and Jelena Kovačević. Designing local orthogonal bases on finite groups. I. Abelian case. *J. Fourier Anal. Appl.*, 6(1):1–23, 2000.

[Chu97] Fan R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997.

[CM91] Ronald R. Coifman and Yves Meyer. Remarques sur l'analyse de Fourier à fenêtre. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(3):259–261, 1991.

[KVV04] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: good, bad and spectral. *J. ACM*, 51(3):497–515 (electronic), 2004.

[MC] M. Maggioni and R. R. Coifman. Diffusion wavelets. *Applied and Computational Harmonic Analysis*.

[MCBS05] M.Maggioni, R. R. Coifman, J. Bremer, and A. Szlam. Diffusion wavelets packets. submitted to Applied and Computational Harmonic Analysis, 2005.

[ROF92] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[SMBC05] A. Szlam, M. Maggioni, J. Bremer, and R. R. Coifman. Diffusion-driven multiscale analysis on manifolds and graphs: top-down and bottom-up constructions. SPIE wavelets XI, 2005.

[Ste70] Elias M. Stein. *Topics in harmonic analysis related to the Littlewood-Paley theory.* Annals of Mathematics Studies, No. 63. Princeton University Press, Princeton, N.J., 1970.

[Tsc02] D. Tschumperle. *PDE's Based Regularization of Multivalued Images and Applications.* PhD thesis, Universite de Nice-Sophia Antipolis, 2002.

[ZMP04] Lihi Zelnik-Manor and P. Perona. Self-tuning spectral clustering. *Eighteenth Annual Conference on Neural Information Processing Systems, (NIPS)*, 2004.