GEOMETRIC HARMONICS

R. R. Coifman and Stéphane Lafon

ABSTRACT. The goal is to efficiently describe functions on a set Γ. In particular, we wish to analyze restrictions of bandlimited functions whose energy is maximized on the set Γ. We form a basis of "prolate-like" functions whose energy is maximized on the set Γ. The corresponding basis functions generalize both spherical harmonics and prolate functions.

In this report, we introduce a family of functions termed *geometric harmonics*. These functions are obtained by diagonalizing a linear operator on a set, and they possess several special properties. The report is organized as follows: we start by presenting the general construction of the geometric harmonics, then we focus on a special case of interest (namely, when the operator to be diagonalized is a projector). We also list the main properties of these functions, and give some examples. We show how geometric harmonics can be used to handle dataset in high dimension, by doing clustering or by parametrizing points on a manifold.

In the sequel, Γ and Ω are two subsets of $\mathbb{R}^n$ such that $\Gamma \subset \Omega$. Let μ be a finite measure on Γ: $\mu(\Gamma) < +\infty$.

1. Definition of the Geometric Harmonics

Let $k : \Omega \times \Omega \longrightarrow \mathbb{R}$ be a kernel (all functions considered in this report are real-valued functions, although the generalization to complex-valued functions is obvious). Suppose that this kernel is positive definite (see appendix A for a definition).

We can view $\{k(x,y), x \in \Gamma, y \in \Gamma\}$ as a a matrix indexed by the points of Γ and since this matrix is positive (and symmetric) so we can think of diagonalizing it. Let's define φ_j, and λ_j $(j \geq 0)$ by

$$\lambda_j \varphi_j(x) = \int_\Gamma k(x,y)\varphi_j(y)d\mu \tag{1}$$

for $x \in \Gamma$. Let's justify the existence of the φ_j's and λ_j's.

Let $\mathcal{H}$ be the reproducing kernel Hilbert space corresponding to k (see appendix B for an explanation) .

EXAMPLE 1. It is shown in appendix B that if $k(x,y) = e^{-\frac{\|x-y\|^2}{t}}$ and $\Omega = \mathbb{R}^n$, then the $\mathcal{H}$ is the space of all functions whose Fourier transform belong to $L^2(\mathbb{R}^n, e^{t\|\xi\|^2}d\xi)$. Likewise, if $k(x,y)$ is the Bessel kernel, then $\mathcal{H}$ is a space of

Last updated on 10/10/2003.

square integrable bandlimited functions. More generally, $\mathcal{H}$ will be a space of functions where the decay of the Fourier transform is imposed (most of their energy concentrated within low frequencies).

Consider the operator $T : L^2(\Gamma, d\mu) \longrightarrow \mathcal{H}$ defined by

$$Tf(x) = \int_\Gamma k(x,y)f(y)d\mu$$

where $x \in \Omega$ and its adjoint $T^* : \mathcal{H} \longrightarrow L^2(\Gamma, d\mu)$ given by

$$T^*g(y) = \langle k(.,y), g(.)\rangle_{\mathcal{H}}$$

where $y \in \Gamma$ and $\langle .,. \rangle_{\mathcal{H}}$ is the inner product in $\mathcal{H}$.

Let's make the technical assumption that

$$x \longmapsto k(x,x) \text{ is } d\mu\text{-essentially bounded on } \Gamma \tag{2}$$

This condition can be weakened for the following to hold, but at least it is easy to check.

Then it can be shown (see appendices D, C and E) that

- $T : L^2(\Gamma, d\mu) \longmapsto \mathcal{H}$ is bounded
- Because of the reproducing property of k, T^* is nothing but the operator $D : \mathcal{H} \longrightarrow L^2(\Gamma, d\mu)$ of restriction of elements of $\mathcal{H}$ onto Γ. This operator is bounded.
- T^*T is Hilbert-Schmidt and therefore can be diagonalized. Because of the observation that $T^* = D$, to obtain the eigenfunctions of T^*T, we diagonalize the kernel k on the set Γ.

The last point above shows that the φ_j's and λ_j's defined in equation (1) exist. Since T^*T is compact, we can order the λ_j's so that they form a non increasing sequence, tending to 0. Moreover, $\{\varphi_j\}$ is an orthogonal sequence in $L^2(\Gamma, d\mu)$.

φ_j is defined on Γ and in fact, if $\lambda_j > 0$, it can be extended to Ω using the following formula:

$$\phi_j(x) = \frac{1}{\lambda_j}\int_\Gamma k(x,y)\varphi_j(y)d\mu \tag{3}$$

This extension process is called Nyström extension and up to the scale factor λ_j, ϕ_j satisfies a generalized mean value theorem. In this report, we show that these extensions can be used to analyse the geometry of sets. The functions ϕ_j are called *geometric harmonics.*

REMARK 1. We have that $T\varphi_j = \lambda_j\phi_j$, therefore $TT^*\phi_j = \phi_j$ and TT^* is self-adjoint, and as a consequence, $\{\phi_j\}$ is an orthogonal sequence in $\mathcal{H}$.

In summary, if one diagonalizes the positive definite kernel k on Γ, one obtains functions $\{\varphi_j\}$ that are orthogonal on Γ for the inner product of $L^2(\Gamma, d\mu)$, and their Nyström extensions, the geometric harmonics $\{\phi_j\}$, are orthogonal on Ω, for the inner product of $\mathcal{H}$.

REMARK 2. Suppose that Γ has a positive Lebesgue measure in Ω, and let B be the orthogonal projector onto $\mathcal{H}$, i.e. $Bf(x) = \langle k(x,.), f(.)\rangle_{\mathcal{H}}$. If D is thought of as an operator from functions defined on Ω into functions defined on Ω (by extending the restrictions with 0), then

$$T = BD$$

and

$$T^*T = DBD$$

since D and B are both projectors, the diagonalization of $T^*T = DBD$ tells us what happens when one successively applies B and D to a function supported on Γ. This is exactly how Slepian *et al* (see [**BellLabs1**]) introduced the prolate spheroidal wave functions in signal processing.

A natural situation where one needs to diagonalize a positive definite kernel on a set is the following: one is given a family $\{e_\xi\}_{\xi \in I}$ of (smooth) functions defined on Ω, and one views their restrictions on Γ as vectors (or points) in a vector space. These restrictions form a redundant set of functions on Γ and one wishes to analyze this set. Suppose one is also given a probability measure p on I (I is possibly non-countable). Then we can view these restrictions as forming a cloud of points, and the probability measure as being a mass distribution on this pointset. To find the principal axes of inertia, we can apply the common analysis technique called Principal Component Analysis. We form the covariance matrix: for x and y in Γ,

$$k(x, y) = \int_{\xi \in I} e_\xi(x)\overline{e_\xi(y)}dp(\xi)$$

and its diagonalization gives the principal directions $\{\phi_j\}_{j \in \mathbb{N}}$. Being a covariance matrix, the kernel $k(x, y)$ is positive definite. In fact the converse is also true, every positive definite kernel can be seen as a covariance matrix: indeed, the diagonalization on Γ yields:

$$k(x, y) = \sum \lambda_j \phi_j(x)\phi_j(y)$$

The motivation of the rest of the report is to analyze the set Γ via a geometric study of certain classes of functions defined on the set.

In the following, the symbol ϕ_j will be used to denote the extensions (functions defined on Ω) while φ_j will denote the restrictions to Γ.

We introduce the following finite dimensional spaces:

$$L^2_\varepsilon(\Gamma, d\mu) = \text{span}\{\varphi_j(x), x \in \Gamma, \text{ such that } \lambda_j > \varepsilon\}$$

and

$$\mathcal{H}_\varepsilon = \text{span}\{\phi_j(x), x \in \Omega, \text{ such that } \lambda_j > \varepsilon\}$$

2. Properties of the geometric harmonics

Now we explore two characteristic properties of these sets of functions

2.1. Double orthogonality.

A remarkable fact on the geometric harmonics is the following:

PROPERTY 1. *The functions $\{\phi_j\}_{\lambda_j > \varepsilon}$ form an orthogonal basis of $\mathcal{H}_\varepsilon$ and their restrictions to Γ form an orthogonal basis of $L^2_\varepsilon(\Gamma, d\mu)$.*

PROOF. By construction, $\{\phi_j\}$ is orthogonal in $L^2_\varepsilon(\Gamma, d\mu)$ since they are the eigenfunctions of T^*T. If $\lambda_j > \varepsilon$, as stated in remark 1, $\{\phi_j\}$ is orthogonal. □

It is also useful to remark that if $\lambda_j > 0$,

$$\begin{aligned}\langle \phi_i, \phi_j \rangle_{\mathcal{H}} &= \frac{1}{\lambda_j} \langle \phi_i, TT^* \phi_j \rangle_{\mathcal{H}} \\ &= \frac{1}{\lambda_j} \langle T^* \phi_i, T^* \phi_j \rangle_{\Gamma} \\ &= \frac{1}{\lambda_j} \langle \varphi_i, \varphi_j \rangle_{\Gamma}\end{aligned}$$

and it follows that

$$\|\varphi_j\|_{\Gamma}^2 = \lambda_j \|\phi_j\|_{\mathcal{H}}^2$$

This property implies that if $f \in \mathcal{H}_\varepsilon$ then

$$f \stackrel{\mathcal{H}_\varepsilon}{=} \sum_{j \in S_\varepsilon} c_j \phi_j$$

and also

$$Df \stackrel{L^2_\varepsilon(\Gamma, d\mu)}{=} \sum_{j \in S_\varepsilon} c_j \varphi_j$$

with

$$\|f\|_{\mathcal{H}}^2 = \sum_{j \in S_\varepsilon} |c_j|^2$$

and

$$\|Df\|_{L^2(\Gamma, d\mu)}^2 = \sum_{j \in S_\varepsilon} |c_j|^2 \lambda_j$$

2.2. Variational optimality.

The geometric harmonics are the optimal solution to the problem of finding the element of $\mathcal{H}$ most concentrated on Γ:

PROPERTY 2. *Assume that ϕ_j is normalized such that $\|\phi_j\|_{\mathcal{H}} = 1$, Then if $\lambda_j > 0$, then ϕ_j is a solution to*

$$\arg\max_{f \in \mathcal{H}} \|Df\|_{\Gamma}^2$$

under the constraints:

$$\|f\|_{\mathcal{H}} = 1 \textit{ and } f \perp \{\phi_0, \phi_1, ..., \phi_{j-1}\} \textit{ in } \mathcal{H}$$

In particular, ϕ_0 is the element of $\mathcal{H}$ with unit energy which best concentrated on Γ.

Note that there is not necessarily uniqueness of the solution since the eigenspaces can be degenerate (if Γ possesses some symmetry).

PROOF. We want to maximize $\langle Df, Df \rangle_{\Gamma} = \langle TT^* f, f \rangle_{\mathcal{H}}$ under the constraints: $\langle f, f \rangle_{\mathcal{H}}^2 = 1$ and $\langle \phi_i, f \rangle_{\mathcal{H}} = 0$, for $0 \leq i \leq j-1$. The Lagrange multipliers technique says that there exist scalars λ and $\alpha_0, ..., \alpha_{j-1}$ such that

$$TT^* f = \lambda f + \alpha_0 \phi_0 + ... + \alpha_{j-1} \phi_{j-1}$$

Taking the inner product in $\mathcal{H}$ with ϕ_i $(0 \leq i \leq j-1)$ yields

$$\alpha_i = \langle TT^* f, \phi_i \rangle_{\mathcal{H}} = \langle f, TT^* \phi_i \rangle_{\mathcal{H}} = \lambda_i \langle f, \phi_i \rangle_{\mathcal{H}} = 0$$

thus $TT^* f = \lambda f$ and taking the inner product with f yields

$$\langle Df, Df \rangle_{\Gamma} = \lambda$$

This quantity will be maximized if $f = \phi_j$ □

3. Examples

3.1. The Prolate Spheroidal Wave Functions.

In [**BellLabs1**], [**BellLabs2**] and [**BellLabs3**], Slepian *et al* introduced the Prolate Spheroidal Wave Functions as the solution of the problem of finding functions "optimally" concentrated in time and in frequency. In their papers, they define $\mathcal{H}^c$ to be the set of functions in $L^2(\mathbb{R}^n)$ whose Fourier transform are compactly supported in the ball centered at the origin and of radius $\frac{c}{2}$. In other words, it is the set of bandlimited square summable functions, with bandwidth equal to $\frac{c}{2}$.

In this case, the kernel k is given by (see appendix F):

$$k_c(x,y) = \left(\frac{c}{2}\right)^{\frac{n}{2}} \frac{J_{\frac{n}{2}}(\pi c\|x-y\|)}{\|x-y\|^{\frac{n}{2}}}$$

where J_ν is the Bessel function of the first kind and order ν. We will refer to this kernel as the Bessel kernel of dimension n.

In the case when n is odd, this kernel can be expressed in terms of the derivatives of the sinc function(see appendix F). For $n = 1$, the kernel is simply given by

$$k_c(x,y) = c\mathrm{sinc}(c(x-y))$$

The eigenfunctions of this kernel on a set Γ of positive Lebesgue measure (in Ω) were termed Prolate Spheroidal Wave Functions. When Γ is embedded in a vector space of higher dimension (i.e. if the codimension $p > 1$), then it can be analyzed using any Bessel kernels of dimension $\geq n - p$, and the eigenfunctions are more generally called (bandlimited) geometric harmonics.

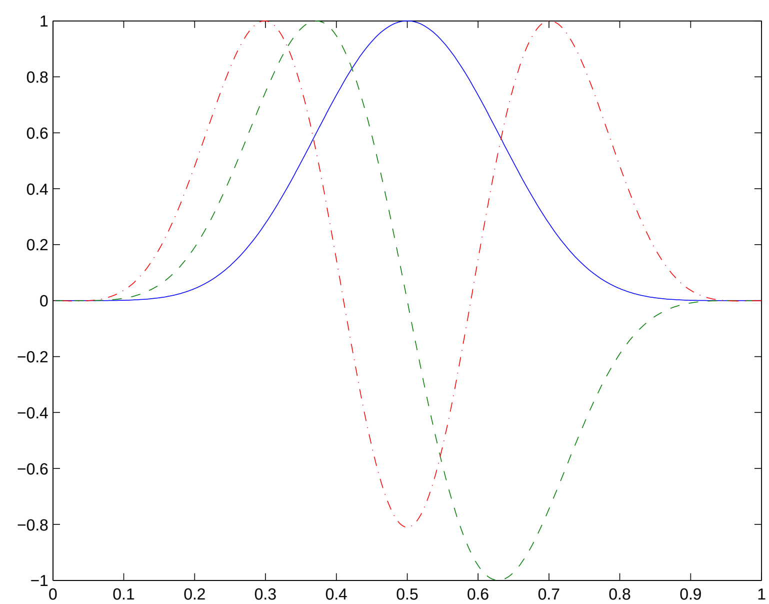

FIGURE 1. Slepian prolates on $\Gamma = [0, 1]$ with the sinc kernel.

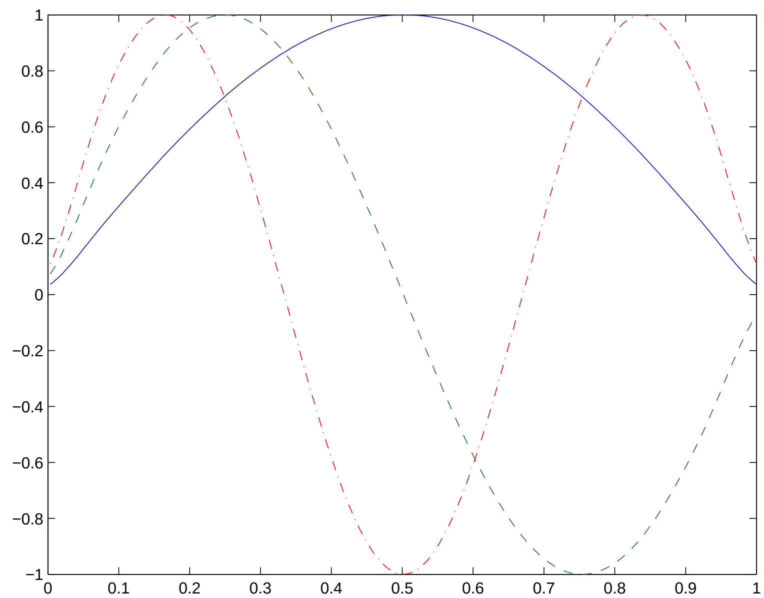

FIGURE 2. Slepian prolates on $\Gamma = [0, 1]$ with the Bessel kernel, $n \geq 2$.

Since $\mathcal{H}^c$ is a set of bandlimited functions, its elements are analytic and any set Γ containing an open ball is a set of uniqueness. Therefore the prolates form an orthonormal basis of $\mathcal{H}^c$ and their restriction to Γ form an orthogonal basis of $L^2(\Gamma, d\mu)$. It can be checked that all eigenvalues are non-zero, and that if the codimension of Γ is $p = 0$ (if Γ contains an open set for instance), then the eigenvalues belong to $(0, 1)$.

Note that by construction, k_c is rotation invariant and as a consequence, if Γ is a circle, then T_c is a convolution operator, and its eigenfunctions are the Fourier basis functions. If Γ is a sphere, then the eigenfunctions are the spherical harmonics.

3.2. Geometric harmonics based on Fourier series: the discrete prolates.

Consider the torus $\Omega = \mathbb{T} = \mathbb{R}/\mathbb{Z}$ and let $q \geq 0$. Let

$$\mathcal{H} = \text{span}\{e^{2i\pi jx}\}_{|j| \leq q}$$

The corresponding kernel is the Dirichlet kernel:

$$k(x, y) = D_q(x - y) = \sum_{j=-q}^{q} e^{2i\pi j(x-y)} = \frac{\sin(\pi(x-y)(2q-1))}{\sin(\pi(x-y))}$$

Take $\Gamma = (-\frac{a}{2}; \frac{a}{2})$ with $0 < a \leq 1$. Then Γ is a set of uniqueness. Writing

$$\psi(x) = \sum_{j=-q}^{q} c_j e^{2i\pi jx}$$

$(c_{-q}, c_{-q+1}, ..., c_q)^T$ is an eigenvector of the Toeplitz matrix

$$\{a\text{sinc}(\pi a(i-j))\}_{|i|,|j| \leq q}$$

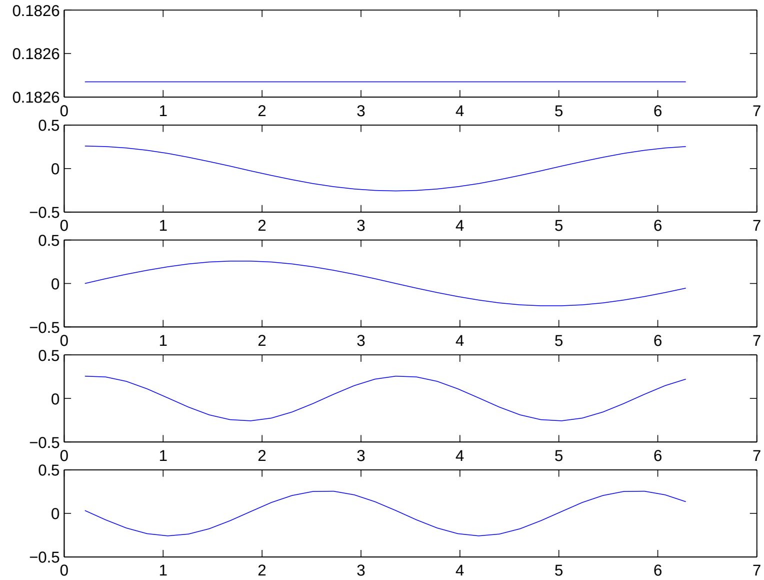

FIGURE 3. Slepian prolates on the circle: we obtain the Fourier basis.

These eigenvectors correspond to the discrete prolates (see [**BellLabs5**]).

3.3. A simple example.

Let

$$\Omega = \bigsqcup_{j\in\mathbb{Z}} A_j$$

be a partition of Ω, and let

$$k(x,y) = \sum_{j\in\mathbb{Z}} \chi_{A_j}(x)\chi_{A_j}(y)$$

The corresponding Hilbert space $\mathcal{H}$ is the set of functions that are constant on the A_j and the projector B averages functions of $L^2(\Omega)$ over the sets A_j. In this case, the geometric harmonics are all the functions χ_{A_j}. Indeed,

$$T\chi_{A_j} = \frac{|\Gamma \cap A_j|}{|A_j|}\chi_{A_j}$$

where $|A_j \cap \Gamma|$ and $|A_j|$ are the volumes of $A_j \cap \Gamma$ and A_j.

0 belongs to the spectrum of DT, and the geometric harmonics are all the functions χ_{A_j} for which $|A_j \cap \Gamma| \neq 0$.

More generally, examples of this sort can be constructed by considering not only a single partition of Ω but a filtration.

3.4. Geometric harmonics based on wavelet multiresolutions.

Let $\{V_j\}_{j\in\mathbb{Z}}$ be a wavelet multiresolution in $L^2(\Omega)$. There are two ways to obtain geometric harmonics from this construction:

- *Using the scaling function* Φ*:* inspired from the previous example, we take $\mathcal{H} = V_j$. When the wavelets are compactly supported, 0 is an eigenvalue of DT. On the contrary, if we choose Meyer wavelets (analytic), then if Γ contains an open ball, it is a set of uniqueness. However, Meyer wavelets (and the corresponding scaling functions) decay exponentially, and infinitely many eigenvalues λ_j will be exponentially small.
- *Using the wavelet* Ψ*:* Instead of using V_j, we can use the wavelet spaces W_j. We take $\mathcal{H} = \bigcup_{j\in J} W_j$. Again, when the wavelets are compactly supported, 0 is an eigenvalue of DT. The same remark concerning Meyer wavelets holds.

4. Bandlimited geometric harmonics

In this section, we study the geometric harmonics obtained as eigenfunctions of the Bessel kernels (when the domain Γ has positive Lebesgue measure, these functions are commonly refered to as prolates). We essentially prove that as the bandwidth tends to infinity, the geometric harmonics become localized on the set and allow a separation of its different connected components (up to some resolution). We also mention a property on the oscillating behavior (number of zeros) of the geometric harmonics.

4.1. Localization property.

We now show that a first possible application of the geometric harmonics: *clustering.*

Suppose that the dimension is $n = 1$ and that Γ consists of two disjoint segments in $\mathbb{R}$:

$$\Gamma = \Gamma_1 \sqcup \Gamma_2$$

and we suppose that the distance between Γ_1 and Γ_2 is $d > 0$. In this case,

$$k_c(x, y) = c\,\mathrm{sinc}(c(x-y))$$

Let f be a smooth function defined on Γ and such that $f_{|\Gamma_2} = 0$. Then for $x \in \Gamma_2$

$$\begin{aligned}
T_c f(x) &= \int_\Gamma c\mathrm{sinc}(c(x-y))f(y)dy \\
&= \int_{\Gamma_1} \sin(\pi c(x-y))\frac{f(y)}{\pi(x-y)}dy \\
&= \sin(\pi cx)\int_{\Gamma_1}\cos(\pi cy)\frac{f(y)}{\pi(x-y)}dy - \cos(\pi cx)\int_{\Gamma_1}\sin(\pi cy)\frac{f(y)}{\pi(x-y)}dy \\
&= \mathcal{O}\left(\frac{1}{cd}\right)
\end{aligned}$$

by integration by part (this is the Riemman-Lebesgue lemma) and using the fact that $|x-y| \geq d$.

This calculation proves that if c is large enough, i.e. if $c \gg \frac{1}{d}$, then for all $x \in \Gamma_2$ $T_c f(x) \simeq 0$. This is an important property: *the space of functions supported on a given connected component of* Γ *is (numerically) left invariant by* T_c, provided that c is sufficiently large. An illustration of this fact is shown on figure 4. The matrix of T_c is approximately a block matrix.

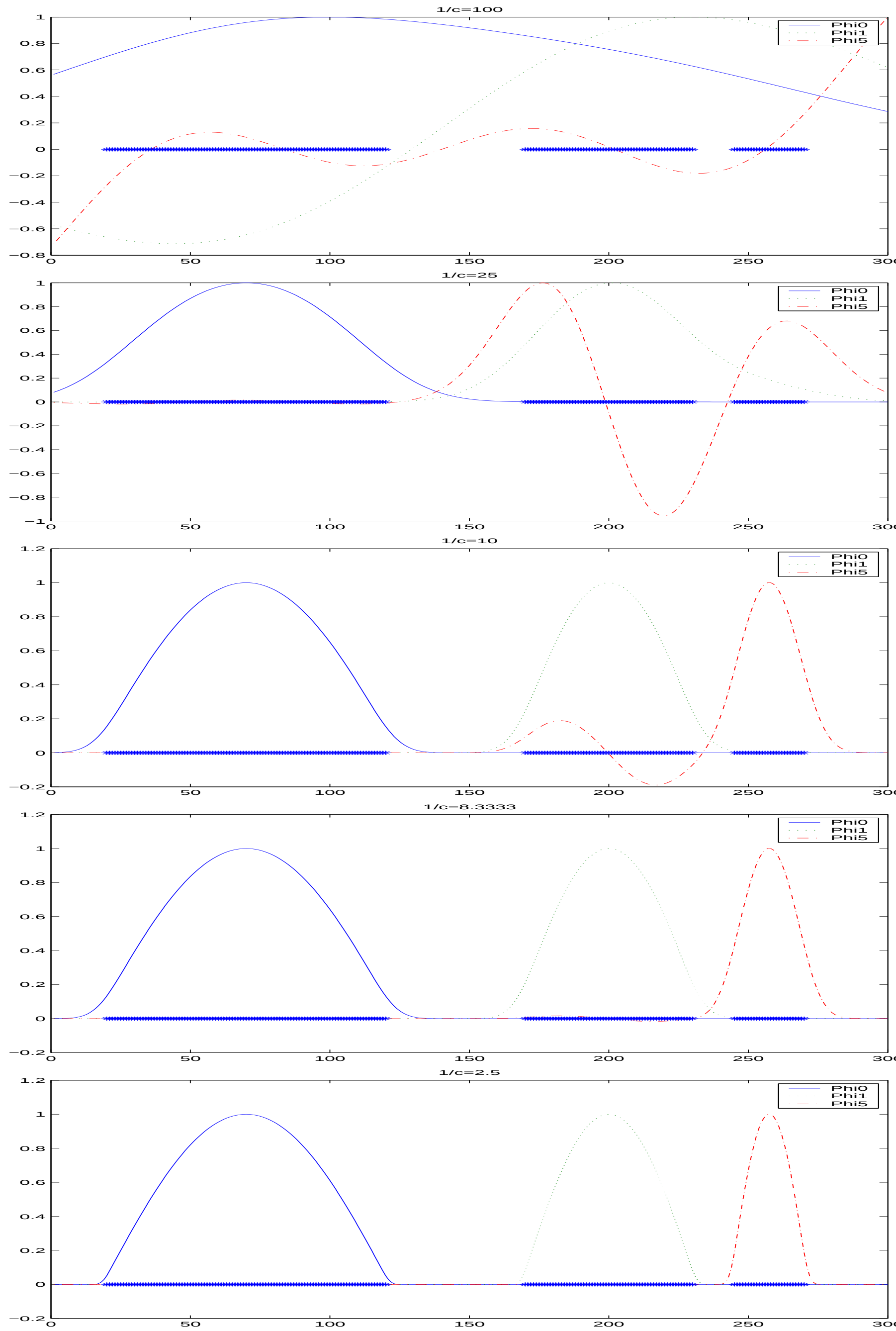

FIGURE 4. Γ is the disjoint union of three disjoint intervals $I_1 \cup I_2 \cup I_3$. The distance between I_1 and I_2 is equal to 70, and the distance between I_2 and I_3 is equal to 10. The figure shows the geometric harmonics ϕ_0, ϕ_1, and ϕ_5 for different values of c. When $\frac{1}{c} = 100$, none of the intervals can be separated. When $\frac{1}{c} = 25 < 70$, intervals I_1 and I_2 are separated but $I_2 \cup I_3$ is still seen as one unique structure. Last, when $\frac{1}{c} = 2.5 < 10$, the intervals I_2 and I_3 are separated.

As a consequence, each geometric harmonic is a function numerically supported on a connected component of Γ. They can therefore be used to identify and separate the different components of a set up to the resolution c. *The geometric harmonics give us the structure of the set.* In particular they can be efficiently used for clustering purposes.

This localization property remains true in higher dimension because then

$$k_c(x,y) = \left(\frac{c}{2}\right)^{\frac{n}{2}} \frac{J_{\frac{n}{2}}(\pi c\|x-y\|)}{\|x-y\|^{\frac{n}{2}}}$$

and as c increases, the oscillations in $J_{\frac{n}{2}}(\pi c\|x-y\|)$ get wild. See figure 5 for an example in dimension 2.

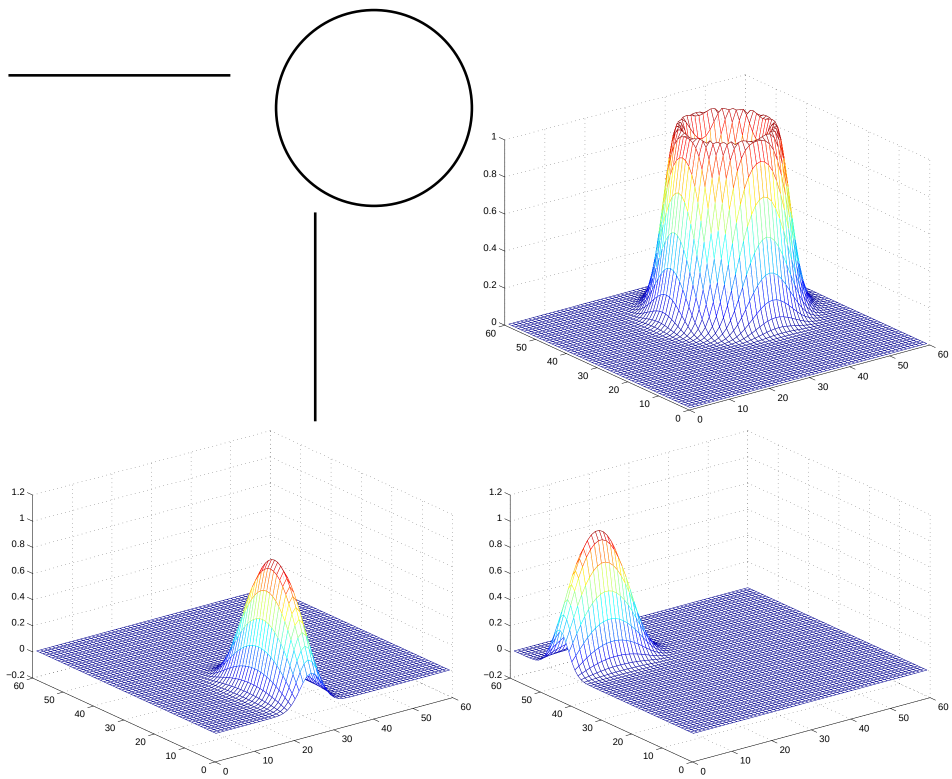

FIGURE 5. Γ is the union of a circle and two segments. We obtain three families of geometric harmonics, each of which living either on the circle or on one of the segments. Above: graphs of the geometric harmonics of order 0 for these three families.

4.2. Number of zeros.

If Γ is an interval and μ is uniform, then ϕ_j has exactly j zeroes. If Γ is a circle, and μ is uniform, then ϕ_j has $2j$ zeros. In fact, for these two examples, the prolates are the eigenfunction of some Sturm-Liouville differential operator. For more general Γ of dimension 1, numerical evidences support the fact that the number of zeros of ϕ_j grows linearly with j. In fact, the Bessel kernel is totally positive, and by Krein's theorem, its eigenfunctions have the property that ϕ_j has j zeros.

5. Set filtering, spectral clustering and link with the laplacian

The localization property of the geometric harmonics make them suitable for clustering: the eigenfunctions will be localized on different structures of the set which can therefore be separated. In fact, the idea of using the eigenfunctions of some operator on a set to find clusters, also called spectral clustering, has been widely used for instance in image segmentation (see [**YWeiss**]). The operator that has naturally been considered is the laplacian operator.

Suppose that Γ is a riemannian manifold, that $x \in \Gamma$ and consider the wave equation on Γ:

$$\frac{\partial^2 u}{\partial t^2} = \Delta u$$

with initial condition $u(y,0) = \delta(y-x)$ and $\frac{\partial u}{\partial t}(y,0) = 0$. We can think of this problem as modeling waves on a membrane (or the surface of a drum) when the membrane is briefly hit at the point x at time $t = 0$.

Then the solution is

$$u(y,t) = \sum_{j \geq 0} \phi_j(x)\phi_j(y)\cos(\lambda_j t)$$

where λ_j and ϕ_j are the eigenvalues and eigenfunctions (here normalized on Γ) of Δ on Γ. This spectral representation can be interpretated in a simple way: the wave at the point y is the superposition of different monochromatic waves corresponding to the frequency λ_j with the amplitude $\phi_j(x)\phi_j(y)$. Therefore the $\sum \phi_j(x)\phi_j(y)$ is a measure of the acoustic interaction of the two points x and y in Γ.

For other operators based on a radial kernel $k(x,y) = h(\|x-y\|)$, it can be shown that, in general, we have the following asymptotic expansion in the scale:

$$T_c = (\alpha + \frac{\beta}{c^2})I + \frac{\gamma}{c^2}\Delta + \mathrm{o}\left(\frac{1}{c^2}\right)$$

where $\alpha = \int_{\mathbb{R}^n} h(\|x\|)dx$, and β and γ depend on the geometry of Γ (curvature, torsion...). Therefore, asymptotically for small scales, diagonalizing T_c is equivalent to diagonalizing the laplacian.

Now if we fix the scale $\frac{1}{c}$ (c is not necessarily large). The kernel $k(x,y)$ can be renormalized in the following way:

$$\widetilde{k}(x,y) = \frac{k(x,y)}{\lambda_0\phi_0(x)\phi_0(y)}$$

and the measure μ can be renormalized as $d\widetilde{\mu} = \phi_0(y)^2 d\mu$ on Γ so that if

$$\widetilde{T}_c g(x) = \int_\Gamma \widetilde{k}(x,y)g(y)d\widetilde{\mu}$$

we have that $\widetilde{T}_c$ is an averaging operator: $\widetilde{T}_c 1 = 1$, and as a consequence, $I - \widetilde{T}_c$ is some laplacian operator on Γ.

The measure $d\widetilde{\mu} = \phi_0(y)^2 d\mu$ is a finite measure on Γ and can be used to filter the set, for instance by selecting all points x such that $\phi_0(x)^2$ is above a certain threshold. Then one can start again on the residual points. This way one obtains a hierarchical description of the data set.

This approach can be generalized as follows: let $\varepsilon > 0$ and

$$S_\varepsilon = \{j \in \mathbb{N} \text{ such that } \lambda_j > \varepsilon\}$$

For $x \in \Gamma$, consider

$$p_\varepsilon^c(x) = \frac{1}{|S_\varepsilon|} \sum_{j \in S_\varepsilon} \frac{\phi_j^2(x)}{\|\phi_j\|^2_{L^2(\Gamma, d\mu)}}$$

and for $x \in \Omega$

$$P_\varepsilon^c(x) = \frac{1}{|S_\varepsilon|} \sum_{j \in S_\varepsilon} \frac{\phi_j^2(x)}{\|\phi_j\|^2_{\mathcal{H}}}$$

Then p_ε^c is a probability density on Γ and P_ε^c is a probability density on Ω. P_ε^c is defined on Ω, and is concentrated on Γ.

The probability density P_ε^c can be used to *low-pass filter the set* Γ. Indeed, if we fix a threshold $0 < \tau < 1$ and the resolution c, then the set

$$\Gamma_\tau^c = \{x \in \Omega \text{ such that } P_\varepsilon^c(x) > \tau \|P_\varepsilon^c\|_\infty\}$$

is a filtered version of Γ at the resolution c. An example of such filtering is given figure 6.

We observe a Gibbs phenomenon that is due to the fact that in the definition of P we use only the prolates with eigenvalue $> \varepsilon$. To avoid this effect, we can add some weights to smoothe the effect of truncation:

$$P(x) = \frac{1}{\sum_{j \in S_\varepsilon} w_j} \sum_{j \in S_\varepsilon} w_j \psi_j^2(x)$$

Note that the oscillations in the Gibbs phenomenon automatically "detects" the boundary $\partial\Gamma$ of Γ.

We illustrate the power of the geometric harmonics for clustering with an example (figure 7). $\Gamma \in \mathbb{R}^3$ consists of the union of the segment $M(t) = (1, 0, t)$ and a helix given by the equation:

$$M(t) = (\cos(t), \sin(t), t) \quad \text{with } t \in [0, 2\pi)$$

We fix $\frac{1}{c}$ to be of the order of the minimum distance between two points in the set. When we diagonalize a Bessel kernel on this set, we observe that the spectrum contains two kinds of (nonzero) eigenvalues: those corresponding to geometric harmonics localized on the helix, and those corresponding to geometric harmonics localized on the segment. We select one of these sets, and form the corresponding probability density $p_\varepsilon^c(y)$, we can filter the set (by a thresholding of p_ε^c, as explained above) and recover the structure of the set. See figure 7.

6. Feature finding and low-dimensional embedding of sets

In this section, we address the question of finding structures (features) in a set. In learning theory, many algorithms rely on one's ability to identify features in the training set, and to use these features to perform tasks such as classification and regression. The first point, finding the features, is crucial as the features are supposed to reveal enough structure in the dataset for, say, the classification to achieve good performance. If one has some a priori information the dataset, one should use it in the definition of the features. For instance, if the dataset is formed of different images, using edge and junction detection tools in the definition of the features might not be such a bad idea.

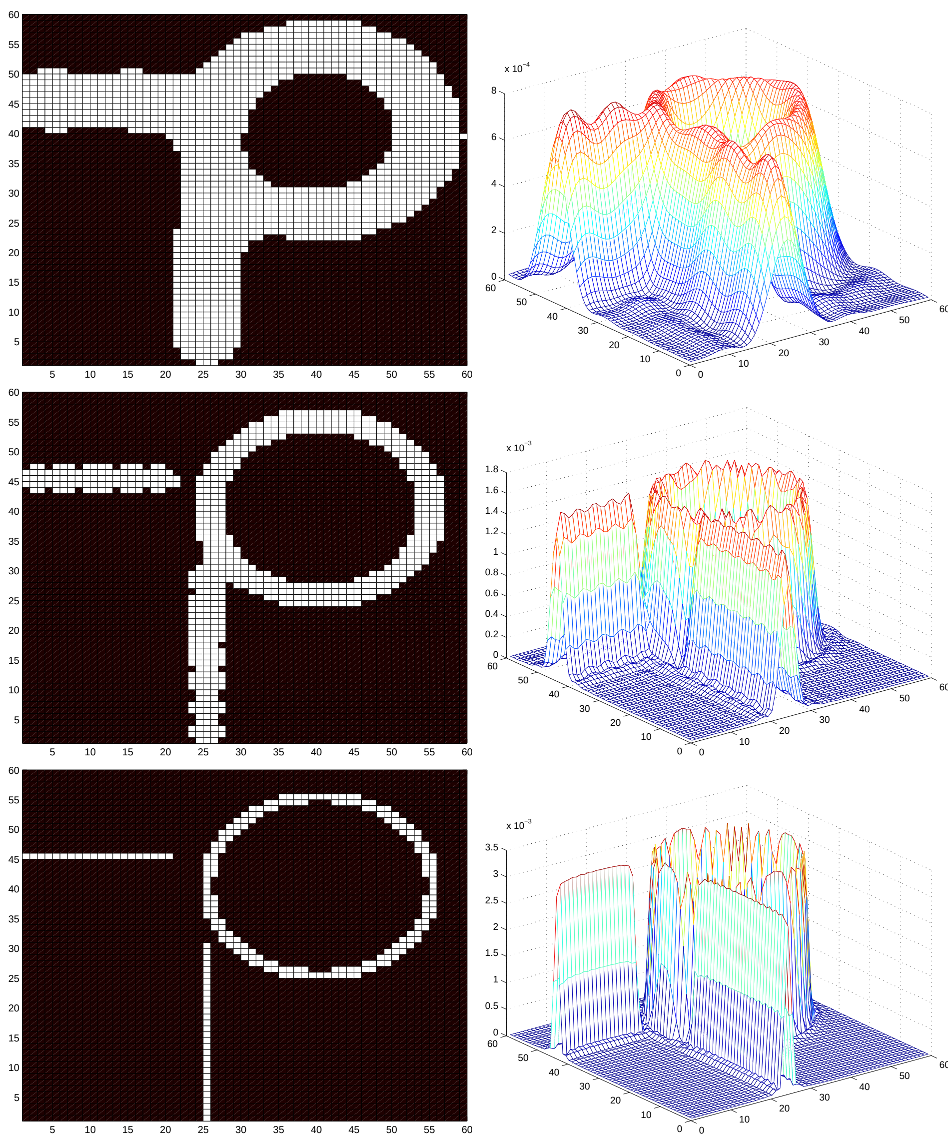

FIGURE 6. Γ is again the union of a circle and two segments. The filtered sets at different resolutions c, and the corresponding density P_ε^c is shown.

However, if no significant a priori knowledge is available, then one has to discover the features from the data (unsupervised learning). The most common technique is the Principle Component Regression (PCR) in which the set is first analyzed via a Principal Component Analysis, then the top eigenvectors (or rather the projections of the points on these vectors) are retained as the features, and the projection on these vectors are used as regressors. So if Γ is approximately a linear manifold, the top eigenvectors provide a "suitable" representation of the points of

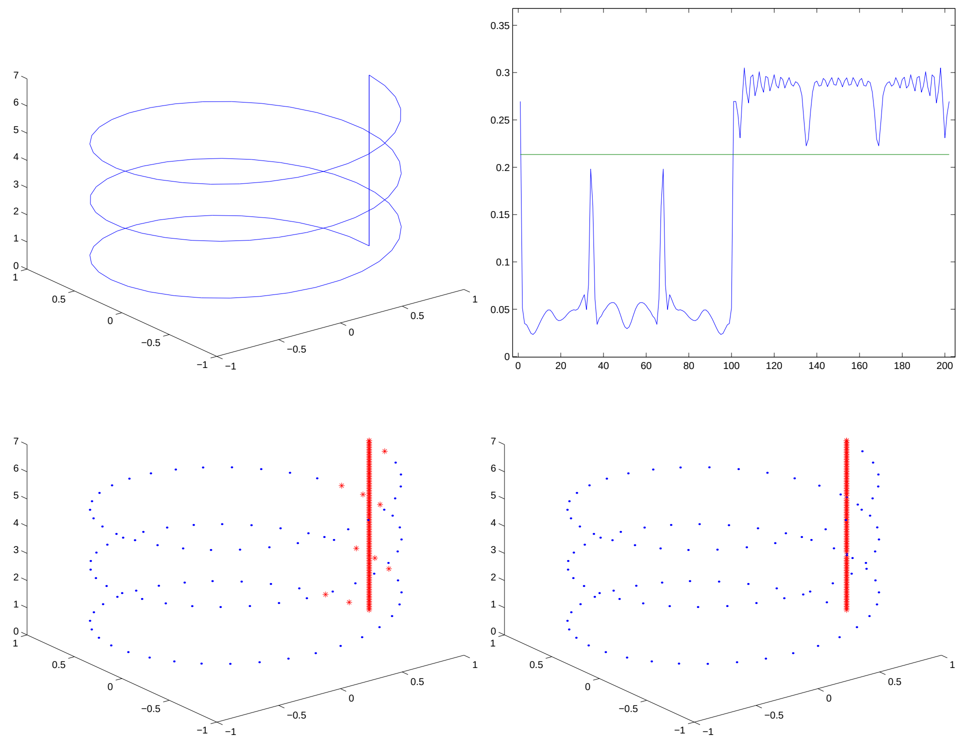

FIGURE 7. Upper Left: Γ is the union of a segment and a helix. Upper right: the probability density p and the threshold. Lower pictures: clustering obtained by thresholding at a low resolution (left) and high resolution (right).

Γ as they form a linear system of coordinates on the set. More generally, if the set Γ is not well represented with a linear system coordinates, then one has to seek non-linear ones. General kernel methods (e.g. SVM) attempt to achieve non-linear descriptions of the dataset by embedding Γ into a non-linear way into a (very) high dimensional space where the points can (hopefully) be described with linear coordinates. In the following, we want to use the geometric harmonics as coordinates on a neighborhood of Γ, and to be able to use these coordinates to discover relevant features on the set.

Suppose that Γ is an abstract set, not necessarily points in the euclidean space $\mathbb{R}^n$, endowed with some distance d. For instance, Γ can be a graph, and d can be obtained by weighting its edges. The question whether (Γ, d) can be globally embedded in a euclidean space $\mathbb{R}^n$ $(1 \leq n \leq +\infty)$ in an isometric way was studied and solved by Schoenberg ([**Schoenberg**]): to answer this question, it suffices to look at the restriction of the Bessel kernels, or the Gaussian kernel to Γ.

Let's form the following vector:

$$\Phi(x) = \begin{pmatrix} \sqrt{\varphi_0}\phi_0(x) \\ \sqrt{\varphi_1}\phi_1(x) \\ \vdots \end{pmatrix}$$

This is an embedding of Γ into $\mathbb{R}^\infty = l^2(\mathbb{N})$. Moreover, by construction,

$$\langle \Phi(x), \Phi(y) \rangle_{l^2} = \sum_{i=0}^{+\infty} \lambda_i \varphi_i(x) \varphi_i(y) = k(x,y) \tag{4}$$

$k(x,y)$ can be thought of as a similarity measure between x and y.

A natural question arises: does this embedding preserve the metric on Γ. In other words is Φ bilipschitz ? For two close points x and y in a strip around Γ, can we find reasonable constants $A < B$ such that :

$$Ad(x,y) \leq \|\Phi(x) - \Phi(y)\|_{l^2} \leq Bd(x,y) \tag{5}$$

holds ?

PROPOSITION 3. *Suppose that $k(x,y) = h(d(x,y)^2)$ where h is $\mathcal{C}^2$. Then the embedding Φ is* locally isometric*:*

$$\lim_{d(x,y) \to 0} \frac{\|\Phi(x) - \Phi(y)\|_{l^2}}{d(x,y)} = C$$

where $C = \sqrt{-h''(0)}$ does not depend on x and y.

PROOF. First note that since by Cauchy-Schwarz, $k(x,y)^2 \leq k(x,x)k(y,y)$, we have $h(d(x-y)^2) \leq h(0)$, and h has a maximum at 0. Therefore, $h'(0) = 0$ and $h''(0) \leq 0$. Now, it can be checked that because of identity (4),

$$\|\Phi(x) - \Phi(y)\|_{l^2}^2 = k(x,x) + k(y,y) - 2k(x,y) = 2[h(0) - h(d(x,y)^2)]$$

and a Taylor expansion yields the result. □

REMARK 4. Note that the assumption that k be radial can be weakened to obtain a bilipschitz map, with Lipschitz bounds independent form the location in the set.

We can define

$$D(x,y) = \sqrt{k(x,x) + k(y,y) - 2k(x,y)}$$

as a new distance on Γ. In fact the proof shows that

$$D(x,y) \simeq \frac{d(x,y)}{1 + d(x,y)}$$

The two distances D and d define the same topology on Γ and they are locally equivalent in the following sense: for x and y such that $d(x,y)$ is much smaller that σ ($\sigma = \frac{1}{c}$ is the scale parameter), $D(x,y) \simeq d(x,y)$, whereas $D(x,y) \simeq 1$ if $d(x,y) \geq \sigma$. Thus D discriminates only between points at (Euclidean) distance less than the size of the support of k, all other points being sent at infinity.

As a similarity measure between x and y, $k(x,y)$ is completely arbitrary since the choice of the kernel may not be related to the set Γ (although the scale parameter c can be adjusted). However, the geometric harmonics are the eigenfunctions of the kernel computed on the set, and together with the eigenvalues provide an analysis of Γ.

Now if we suppose that Γ lies in the neighborhood of some manifold of low dimension in $\mathbb{R}^n$, the embedding Φ is not of much help to recover a parametrization of this manifold. If we use only a small number of geometric harmonics, we can hope to find a system of local coordinates.

EXAMPLE 2. Suppose that Γ is the set of all realizations of the following random process:

$$X_\tau(t) = \begin{cases} 1 & \text{if } |t-\tau| \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

where $|t-\tau|$ is the distance on the torus (circle) $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, $t \in \mathbb{T}$ and τ is uniformly distributed over $\mathbb{T}$.

Γ can be thought of as a curve in the Euclidean space $L^2([0,1], dx)$. Although it was generated in a simple manner by a single parameter τ, it is not very smooth as two points on this curve differ by an orthogonal increment. As a consequence, it has a Hölder $\frac{1}{2}$ regularity:

$$\|X_\tau - X_{\tau'}\| = |\tau - \tau'|^{\frac{1}{2}}$$

This set is rotation invariant and if we take a rotation invariant kernel, like the gaussian $k(x,y) = e^{-\|x-y\|^2}$ then the geometric harmonics are merely the Fourier basis: 1, $\cos 2\pi t$, $\sin 2\pi t$... The choice of $(\cos 2\pi t), \sin 2\pi t)$ as a coordinate system on Γ allows to recover the parameter τ that generated the process (in spite of the poor regularity of the curve).

In the general setting the procedure of finding the coordinates can be carried out through a simple algorithm:

- At each point $x \in \Gamma$, define a neighborhood.
- Find a subset of eigenfunctions such that equation (5) is verified in this neighborhood of x for some bounds A and B. These eigenfunctions define coordinates in this neighbordhood and should be chosen among the ϕ_j's such that $\lambda_j > \varepsilon$
- Agglomerate all related neighborhoods that correspond to the same choice of eigenfunctions.

EXAMPLE 3. Mouse dissection and robustness to noise. Consider a set in $\mathbb{R}^2$ formed of the union of two disks and of a curve (see figure 8). In our experiment we applied the algorithm described above with the gaussian kernel, and we imposed that the ratio of the Lipschitz bounds $\frac{B}{A}$ be less than 5. In other words, the algorithm has decomposed the set into different regions. On each regions, we have a system of coordinates that does not distort small distances too much. We chose $\frac{1}{c}$ to be of the order of 1, wich is the minimum distance between two points in the set. The result is shown on figure 8 (top), and the numbers in the legend correspond to the indices of the geometric harmonics that were selected for a given region. For instance "+ 1 2" means that ϕ_1 and ϕ_2 where selected for the region of all the points which are marked with a '+'.

The body and the ear of the mouse (two disks) is split into 6 domains, each of them being parameterized by 2 geometric harmonics. The tail is divided into 6 domains, each of them being parmetrized by one geometric harmonic. Observe that the correct dimensionality is detected: the body and the ear appear to be one dimensional whereas the tail belong to a one dimensional structure.

Now if we perturb the set with an additional noise, that is if each point $x \in \Gamma$ is changed into the point $x + w(x)$ where $\{w(x)\}_{x\in\Gamma}$ is a gaussian white noise with variance σ^2, then we obtain the results shown on figure 8 (middle and bottom). We have displayed the output of the algorithm for $\sigma = 0.4$ and $\sigma = 1$. This is an

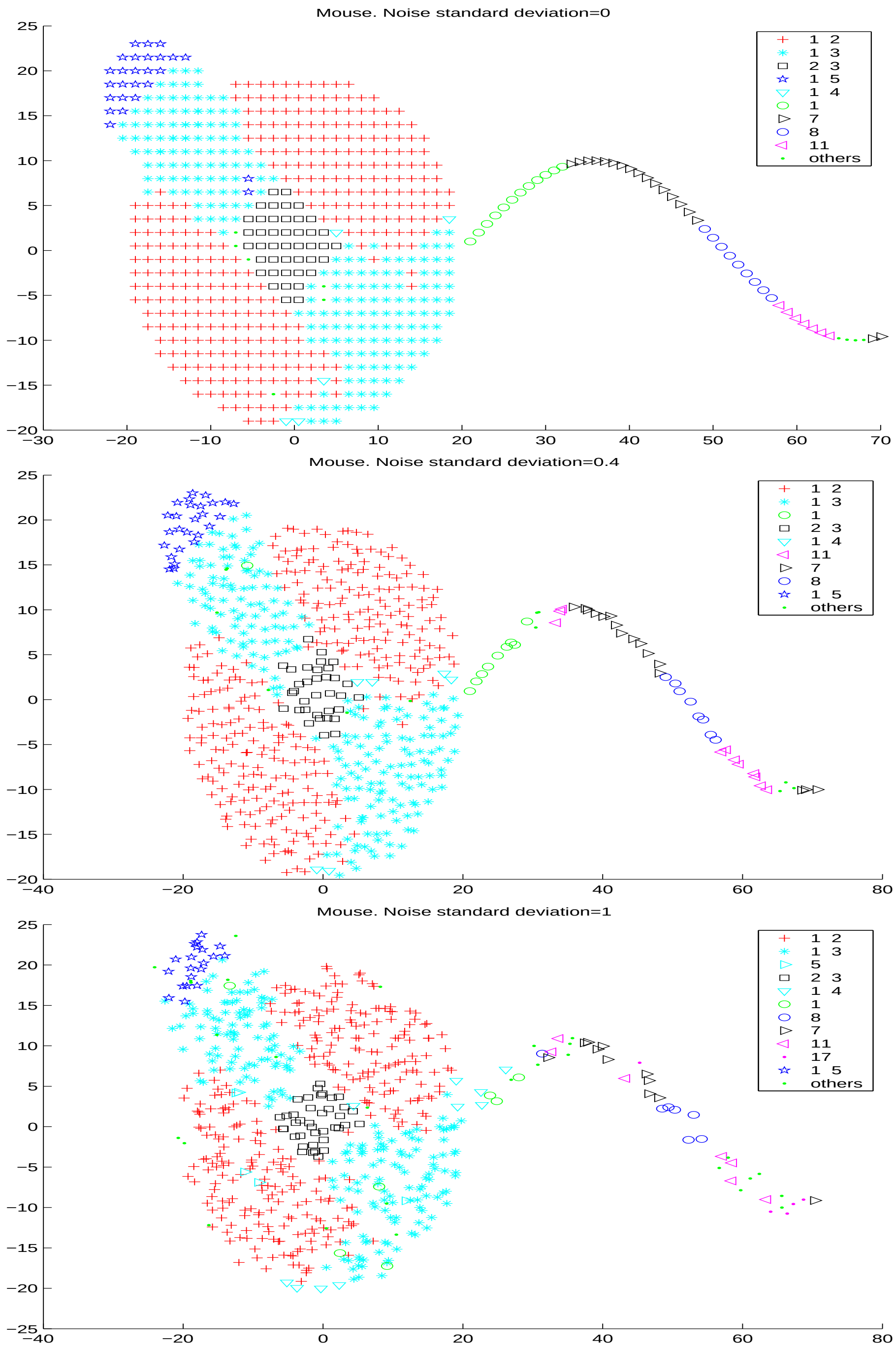

FIGURE 8. Γ is the set formed by points uniformly distributed on the union of two disks and a curve (it's a mouse!). Top: the original mouse set and its decomposition in local patches. The numbers in the legends correspond to the indices of the ϕ_j's selected. Middle: same thing with additive noise with standard deviation=0.4. Bottom: same with standard deviation=1. The decomposition is stable.

important perturbation given that the minimum distance between two points in Γ is 2. These results show that the algorithm is robust to noise.

In fact if the kernel k is smooth, then an additive perturbation on the set Γ is, at the first order, an additive perturbation on the kernel (that is the matrix to be diagonalized). More precisely, if we fix the scale $\frac{1}{c}$ and set if $k_c(x,y) = c^n k(cx, cy)$, then to compute the geometric harmonics, we have to diagonalize the kernel k_c on the set $\{x + w(x), x \in \Gamma\}$. Now,

$$\begin{aligned} k_c(x+w(x), y+w(y)) &= c^n \left[k(c(x+w(x)), c(y+w(y))\right] \\ &\simeq c^n \left[k(cx,cy) + cw(x)\frac{\partial k}{\partial x}(cx,cy) + cw(y)\frac{\partial k}{\partial y}(cx,cy)\right] \\ &\simeq k_c(x,y) + cw(x)\frac{\partial k}{\partial x}(cx,cy) + cw(y)\frac{\partial k}{\partial y}(cx,cy) \end{aligned}$$

Therefore the matrix to be diagonlized is approximately the orignal matrix plus a perturbation. The relative amplitude of this perturbation is of the order of :

$$\left|\frac{k_c(x+w(x), y+w(y)) - k_c(x,y)}{k_c(x,y)}\right| \le c\left(\alpha|w(x)| + \beta|w(y)|\right)$$

where α and β are constants depending on k only. This approximation holds if $c(|w(x)| + |w(y)|) \ll 1$.

We can now invoke all the machinery of matrix perturbation: if one additively perturbs a symmetric matrix A with a matrix E, $\widetilde{A} = A + E$, then there are known results about the spectrum and eigenfunctions of $\widetilde{A}$. For instance Weyl's theorem states that for all j,

$$|\widetilde{\lambda}_j - \lambda_j| \le \|E\|_2$$

Likewise, Mirsky's theorem says that

$$\sqrt{\sum_j (\widetilde{\lambda}_j - \lambda_j)^2} \le \|E\|_F$$

where $\|E\|_F$ is the Frobenius norm of E. These theorems show that the spectrum is robust with respect to noise. Eigenfunctions, on the contrary, are not robust to noise as some perturbation can make an eigenspace degenerate. However, the projector on the space $\{\phi_j, \lambda_j > \varepsilon\}$ is robust to noise when the variance of the noice is much smaller than ε.

In the example of th mouse, we observe that the decomposition of the set into patches as well as the systems of coordinate on these patches remain roughly the same when noise is added, and the algorithm is stable.

EXAMPLE 4. Black disk on white background.

In this example, we analyze a simple image consisting of a black disk on a white backdround. The image is 50x50 pixels. To each pixel we associate its 9x9 neighborhood in the image and we see this neighborhood as a point in $\mathbb{R}^{81}$. This way we have a data set consisting of points in $\mathbb{R}^{81}$ corresponding to all overlapping 9x9 windows containing only black and white pixels separated by an edge. More precisely, the data set is composed of several instances of the totally black window, several instances of the totally white windows, and windows contaning black and white pixels separated by an edge. These windows being quite small, there are essentially two parameters governing the data set: the first one is the proportion

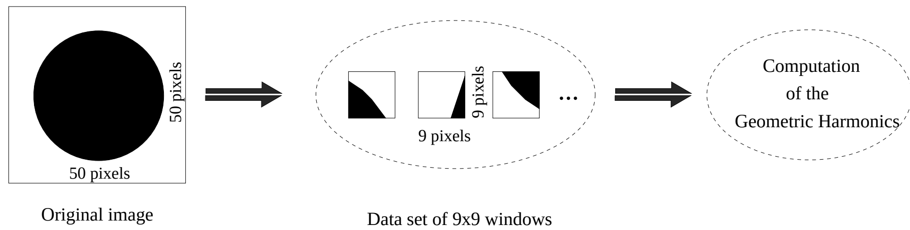

FIGURE 9. The data set is obtained by splitting the image into 9x9 overlapping windows. It is then analyzed via the computation of the geometric harmonics.

of black pixels (or equivalently the offset ot the edge with respect to the upper left corner of the window), and the second one is the orientation of the edge. Note that the dataset has regularity Hölder $\frac{1}{2}$ for the Euclidean norm and as in example 2. For an appropriate choice of the resolution c, the geometric harmonics recover the two parameters. The whole processing of the data is displayed on figure 9.

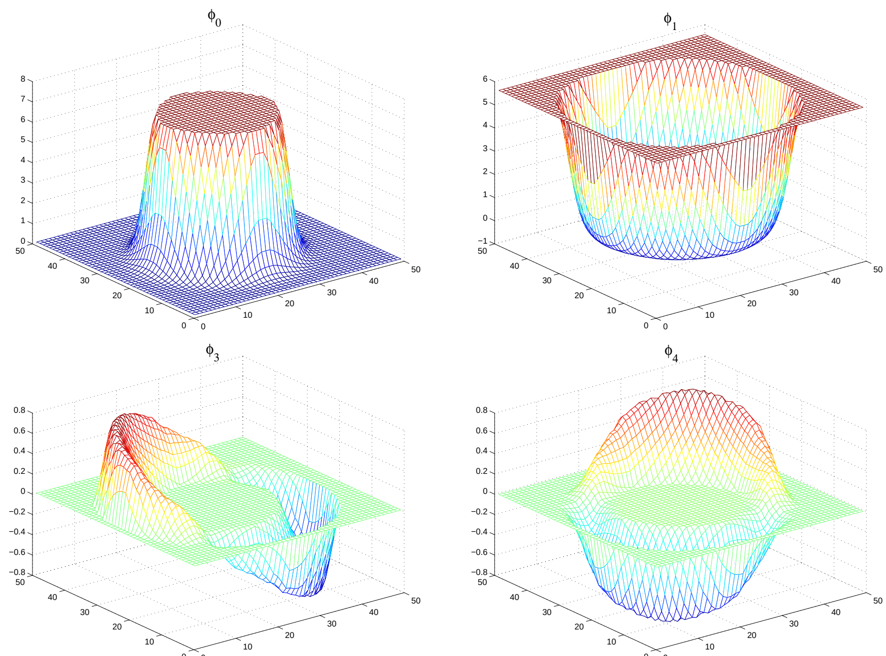

FIGURE 10. ϕ_0, ϕ_1, ϕ_3 and ϕ_4 functions of the pixel position in the image.

We fix $\frac{1}{c}$ to be of the order of the minimum (non zero) distance between two points in the set and we compute the geometric harmonics for this set. The first eigenfunctions are shown on figure 10.

These graphs show that ϕ_0 captures the first parameter (density of black pixel in a window), and so does ϕ_1. While ϕ_0 is mostly concentrated on points corresponding to totally white windows, ϕ_1 is essentially supported by totally black windows.

These functions are not symmetric as the number of totally black windows is greater than the number of totally white windows.

ϕ_3 and ϕ_4 capture the orientation parameter. It is clear that the whole dataset can be parametrized by (ϕ_1, ϕ_2) on one half of the set, and by (ϕ_1, ϕ_4) on the other half. It is even clearer if we take a look at figure 3, where the data set is represented in the coordinates (ϕ_1, ϕ_3, ϕ_4). The points appear to belong to a 2D surface of revolution.

To illustrate the fact that (ϕ_1, ϕ_4) capture the proportion of black pixels and the orientation of the edges, we have represented some windows corresponding to some locations in the plane (ϕ_1^2, ϕ_4) on figure 13. Actually, we have plotted half of the data set, corresponding to the points x where $(\phi_1(x), \phi_4(x))$ provides a bijective parametrization. We observe that as ϕ_1 increases, the proportion of black pixels increases as well. Likewise, the orientation of the edges varies with ϕ_4.

7. Diffusion metric

In the previous section, we have introduced a distance on Γ defined by

$$D^2(x, y) = k(x, x) + k(y, y) - 2k(x, y)$$

or equivalently

$$D^2(x, y) = \sum_{i=0}^{+\infty} \lambda_i (\varphi_i(x) - \varphi_i(y))^2 = \|\Phi(x) - \Phi(y)\|^2$$

We now suppose that the kernel k is rotation-invariant, that is $k(x, y) = h(d(x, y)^2)$ with $h'(0) \neq 0$. Let σ denote the scale parameter and we define $\varepsilon = \sigma^2$.

$$T_\varepsilon f(x) = \int_\Gamma h\left(\frac{d(x, y)^2}{\varepsilon}\right) f(y) d\mu$$

As already mentionned, the kernel can be renormalized so that the corresponding operator is an averaging operator A_ε (in fact there are various ways to normalize the kernel):

$$A_\varepsilon f(x) = \int_\Gamma \frac{h\left(\frac{d(x,y)^2}{\varepsilon}\right)}{\lambda_0 \varphi_0(x) \varphi_0(y)} f(y) \varphi_0(y)^2 d\mu$$

This averaging operator naturally defines a Laplace operator as follows:

$$\Delta_\varepsilon = \frac{A_\varepsilon - I}{\varepsilon}$$

A_ε and Δ_ε correspond to a diffusion on Γ.

It can be shown that there exists a limit operator:

$$\Delta_\varepsilon = \Delta_0 + \varepsilon^{\frac{1}{2}} R_\varepsilon$$

where R_ε is bounded on any space of bandlimited functions. When Γ is a manifold, Δ_0 is the Laplace-Beltrami operator.

From this we can deduce that

$$A_\varepsilon^{\frac{t}{\varepsilon}} = (I + \varepsilon A_\varepsilon)^{\frac{t}{\varepsilon}} \simeq (I + \varepsilon \Delta_0)^{\frac{t}{\varepsilon}} \longrightarrow e^{t\Delta_0}$$

When Γ is a manifold, the kernel $k_t(x, y)$ of A_ε represents the amount of particles (or heat) diffused from x to y at time t. When Γ is a graph, $k_t(x, y)$ is a weighted average over all paths connecting x and y. This similarity measure, and the corresponding metric, provides much information on the geometry of the data.

FIGURE 11. A RGB plot of the image where red=ϕ_1, green=ϕ_3, blue=ϕ_4

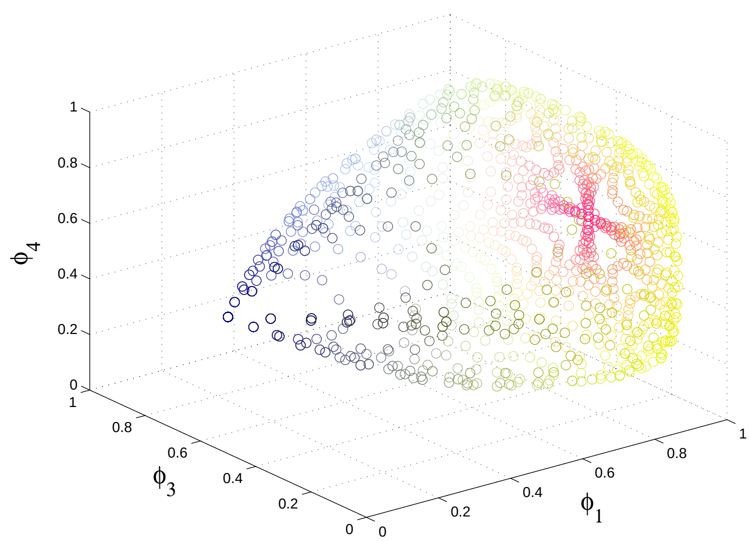

FIGURE 12. Plot of the data in the coordinates (ϕ_1, ϕ_3, ϕ_4). We obtain a surface of revolution. The point $(0, 0, 0)$ corresponds to the totally white windows, while $(1, 0, 0)$ corresponds to the totally black windows.

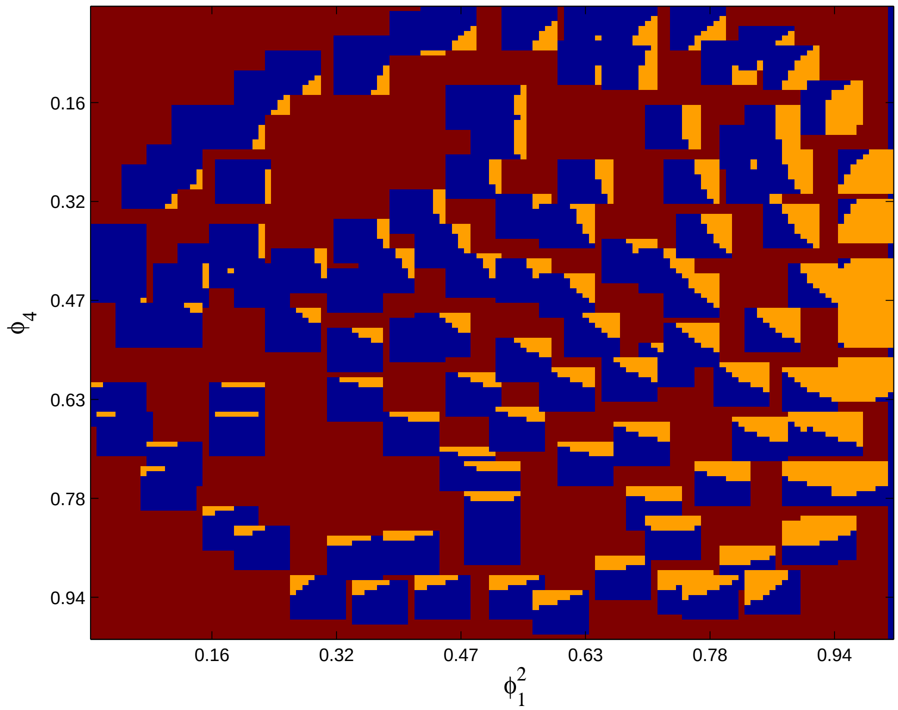

FIGURE 13. Some windows of the data set, in the plane (ϕ_1^2, ϕ_4).

The kernel k_t is thus an approximation of the diffusion kernel defined by $e^{-t\Delta_0}$ and the eigenfunctions ψ_i are approximations of the eigenfunctions of this diffusion kernel with Neuman boundary conditions.

We have the identity:

$$k_t(x,y) = \sum_{i \geq 0} \nu_i^t \psi_i(x)\psi_i(y)$$

where

$$\psi_i(x) = \frac{\varphi_i(x)}{\varphi_0(x)}$$

and

$$\nu_i = \left(\frac{\lambda_i}{\lambda_0}\right)^{\frac{1}{\varepsilon}}$$

It is clear that only a few terms are needed in the sum as $0 \leq \frac{\lambda_i}{\lambda_0} \leq 1$. The corresponding eigenfunctions define a low dimensional embedding which is close to an isometry with respect to the diffusion metric:

$$d_t(x,y) = \sum_{\nu_i^t > 1-\delta} \nu_i^t (\psi_i(x) - \psi_i(y))^2$$

where the equality holds up to exponentially small terms.

Therefore we come to the interesting conclusion that the diffusion on Γ can be computed very easily in the embedding space, namely by

- mapping the data points into a low dimensional space through

$$x \longmapsto \{\psi_i(x)\}$$

- computing the distances using the family of weighted Euclidean distances, where the weights are the given by $\{\nu_i^t\}$.

Various illustrations of this observation are shown on figures 14 and 15.

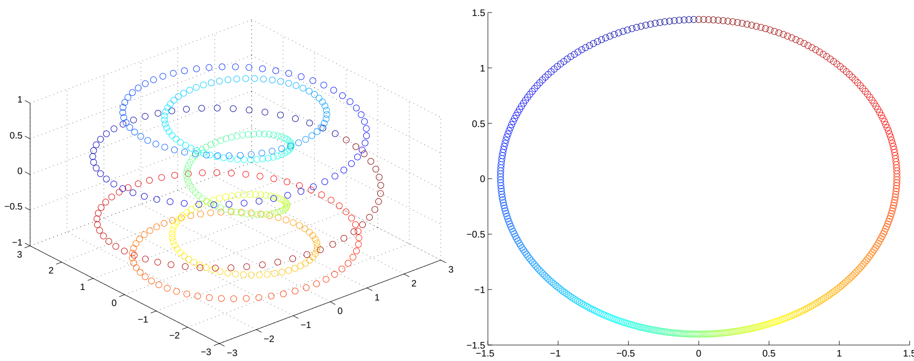

FIGURE 14. The spring curve (left) in $\mathbb{R}^3$ is embedded into $\mathbb{R}^2$ as a circle (right).

The relation between the spectral decomposition of the kernel and the geometry can be explained through the following observations. As already mentionned, to synthesize the kernel k_t, one only needs a finite number of eigenfunctions. More precisely, to reconstruct the bump $k_t(x, y)$ centered at x and of width $\sqrt{t}$, the eigenfunctions needed are those for which ν_i^t exceeds a certain threshold. As t increases, this number gets smaller and the corresponding eigenfunctions have larger support in $\mathbb{R}^d$. This observation, which corresponds to a version of the Heisenberg principle, show how the spectral decomposition of the kernel provide a multiscale analysis of the set.

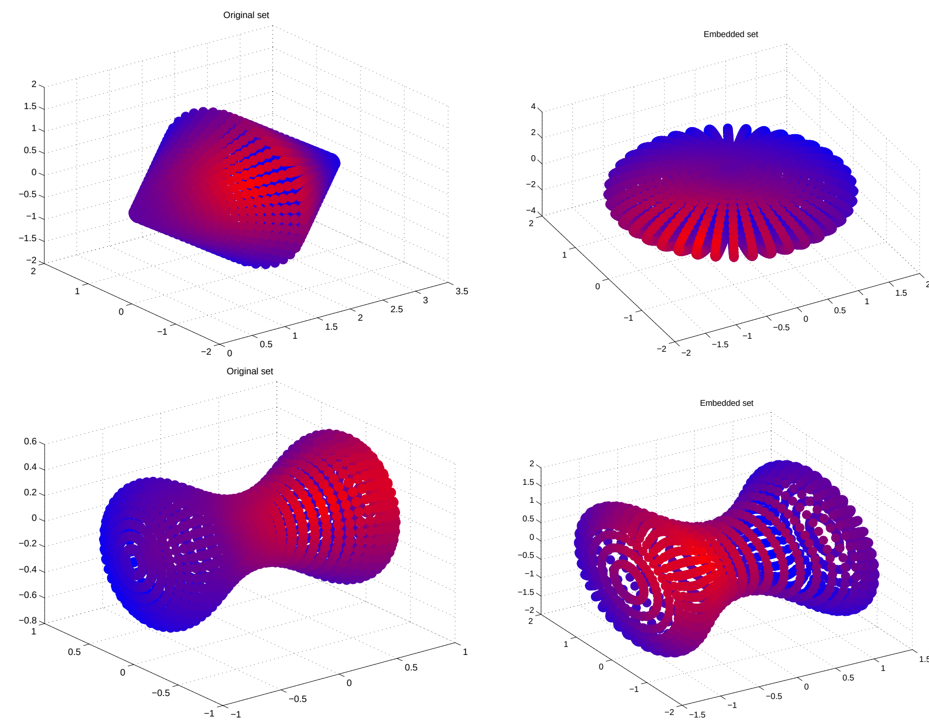

FIGURE 15. Original sets (left) and their embeddings (right). The colors represent the distance to a given point on Γ (red=close, blue=far)

Appendix A. Positive definite kernels

Bochner (see [**Bochner**]) defines (real) positive definite kernels as follows:

DEFINITION 1 (Positive definite kernel). A non zero function $k : \Omega \times \Omega \longrightarrow \mathbb{R}$ is said to be a *definite positive kernel on* Ω if for all $m \geq 1$ and all choices of points $\{x_1, x_2, ..., x_m\}$ in Ω, the matrix $\{k(x_i, x_j)\}$ is positive semidefinite.

Note that the matrix is only semidefinite, not necessarily definite. Moreover it can be proven that a positive definite kernel is necessarily symmetric, that is $k(x, y) = k(y, x)$.

An interesting subclass of positive kernels is given by functions of the form $k(x-y)$ and are obtained by taking the Fourier transform of finite positive measures on $\mathbb{R}^n$ (Bochner's theorem). Among them we find the gaussian kernel, and the Bessel kernels that have the particularity of having a radial Fourier transform, that is these kernels are of the form $k(\|x - y\|)$.

Appendix B. Reproducing kernel Hilbert spaces

A Hilbert space $\mathcal{H}$ of functions defined on (Ω, dx) is said to be a reproducing kernel Hilbert space with kernel k if

- for almost every x in Ω, $k(x, .) \in \mathcal{H}$
- for almost every x in Ω and all $f \in \mathcal{H}$, $\langle k(x, .), f\rangle_{\mathcal{H}} = f(x)$

where $\langle .,. \rangle_{\mathcal{H}}$ denotes the inner product in $\mathcal{H}$. It is easy to verify that any reproducing kernel is a positive definite kernel. In fact the concepts of reproducing kernels and positive definite kernels are identical since if k is a positive definite kernel, one can construct a Hilbert space $\mathcal{H}$ in which k is the reproducing kernel. For more information about this construction process, see [**Aronszajn**].

Since all this might seem abstract, let's give an example. Suppose that $k(x,y) = p(x-y)$ and that $\Omega = \mathbb{R}^n$. Then by Bochner's theorem we know that p is the Fourier transform of a finite positive measure, and for simplicity we will assume that it is of the form $\widehat{p}(\xi)d\xi$. Now consider all signed measures ν verifying

$$\int_\Gamma \int_\Gamma p(x-y) d\nu d\nu < +\infty \tag{6}$$

and define the space of convolutions of all such measures with p:

$$\mathcal{H} = \left\{ f \text{ of the form } f(x) = \int_{\mathbb{R}^n} p(x-y)d\nu \text{ with } \nu \text{ verifying condition (6)} \right\}$$

If we endow this space with the following inner product:

$$\langle f, g \rangle_{\mathcal{H}} = \int_\Gamma \int_\Gamma p(x-y) d\nu_1 d\nu_2$$

where $f(x) = p * d\nu_1$ and $g(x) = p * d\nu_2$, we obtain a Hilbert space and it can be verified that $k(x,y) = p(x-y) = p * \delta_x(y)$ is the reproducing kernel of this Hilbert space.

To understand what this space looks like, we can use the Fourier transform to characterize it. We thus obtain that

$$\langle f, g \rangle_{\mathcal{H}} = \int_{\widehat{p}(\xi)>0} \widehat{f}(\xi)\overline{\widehat{g}(\xi)} \frac{d\xi}{\widehat{p}(\xi)}$$

and this shows that $\mathcal{H}$ is nothing but the image of a weighted L^2 by the Fourier transform. Remember that from Bochner's theorem, $\widehat{p}(\xi) \geq 0$ and is integrable, and therefore the frequency content of elements of $\mathcal{H}$ is concentrated around low frequencies (the weight $\frac{1}{\widehat{p}}$ penalizes high frequencies.

In particular, if $k(x,y) = e^{\|x-y\|^2}$, then $\mathcal{H} = \mathcal{F}^{-1}\left(L^2(\mathbb{R}^n, \beta e^{\alpha \|\xi\|^2})\right)$. Likewise, if $n = 1$ and $k(x,y) = e^{|x-y|}$ then $\mathcal{H}$ is the Sobolev space commonly denoted by W_1^2. Last, if $\widehat{p}$ is the indicator of some set, then $\mathcal{H}$ is the space of all square integrable bandlimited functions (i.e. that have frequencies in this set).

Appendix C. $T^* = D : \mathcal{H} \longrightarrow L^2(\Gamma, d\mu)$ is bounded

Assume that $f \in \mathcal{H}$. Then

$$\begin{aligned} |f(y)| &= |\langle k(y,.), f(.) \rangle_{\mathcal{H}}| \\ &\leq \|k(y,.)\|_{\mathcal{H}} \|f\|_{\mathcal{H}} \quad \text{(Cauchy-Schwarz)} \\ &\leq \sqrt{k(y,y)} \|f\|_{\mathcal{H}} \end{aligned}$$

Thus we can conclude:

$$\int_\Gamma |f(y)|^2 d\mu \leq \|f\|_{\mathcal{H}}^2 \mu(\Gamma) \operatorname{ess\,sup}_{y \in \Gamma} k(y,y)$$

Appendix D. $T : L^2(\Gamma, d\mu) \longrightarrow \mathcal{H}$ is bounded

If $f \in L^2(\Gamma, d\mu)$, then

$$\begin{aligned}
\|Tf\|^2_{\mathcal{H}} &= \left\langle \int_\Gamma k(.,y_1)f(y_1)d\mu(y_1), \int_\Gamma k(.,y_2)f(y_2)d\mu(y_2) \right\rangle_{\mathcal{H}} \\
&= \int_\Gamma \int_\Gamma \overline{f(y_1)} f(y_2) \langle k(.,y_1), k(.,y_2)\rangle_{\mathcal{H}} d\mu(y_1)d\mu(y_2) \\
&= \int_\Gamma \int_\Gamma \overline{f(y_1)} f(y_2) k(y_1,y_2) d\mu(y_1)d\mu(y_2)
\end{aligned}$$

Since k is the reproducing kernel of $\mathcal{H}$

$$\begin{aligned}
|k(y_1,y_2)|^2 &= |\langle k(y_1,.), k(y_2,.)\rangle_{\mathcal{H}}|^2 \\
&\leq \|k(y_1,.)\|^2 \|k(y_2,.)\|^2 \quad \text{(Cauchy-Schwarz)} \\
&\leq k(y_1,y_1)k(y_2,y_2)
\end{aligned}$$

therefore, since $x \mapsto k(x,x)$ is $d\mu$-essentially bounded on Γ, then

$$\begin{aligned}
\|Tf\|^2_{\mathcal{H}} &\leq \int_\Gamma \int_\Gamma |f(y_1)f(y_2)k(y_1,y_2)| d\mu(y_1)d\mu(y_2) \\
&\leq \int_\Gamma \int_\Gamma |f(y_1)f(y_2)\sqrt{k(y_1,y_1)k(y_2,y_2)}| d\mu(y_1)d\mu(y_2) \\
&\leq \left(\int_\Gamma f(y)\sqrt{k(y,y)} d\mu(y) \right)^2 \\
&\leq \left(\int_\Gamma |f(y)|^2 d\mu \right) \left(\int_\Gamma k(y,y) d\mu \right) \quad \text{by Cauchy-Schwarz} \\
&\leq \|f\|^2_{L^2(\Gamma,d\mu)} \mu(\Gamma) \operatorname*{ess\,sup}_{y\in\Gamma} k(y,y)
\end{aligned}$$

Appendix E. $T^*T = DT : L^2(\Gamma, d\mu) \longrightarrow L^2(\Gamma, d\mu)$ is Hilbert-Schmidt

Here we consider $DT : L^2(\Gamma, d\mu) \longrightarrow L^2(\Gamma, d\mu)$, given by

$$DTf(x) = \int_\Gamma k(x,y)f(y)d\mu$$

and we prove that it is Hilbert-Schmidt. The fact that it is self-adjoint is immediate. Let's prove that

$$\int_\Gamma \int_\Gamma |k(x,y)|^2 d\mu(y)d\mu(x) < +\infty$$

Thus

$$\begin{aligned}
\int_\Gamma \int_\Gamma |k(x,y)|^2 d\mu(y)d\mu(x) &\leq \int_\Gamma \int_\Gamma k(x,x)k(y,y) d\mu(y)d\mu(x) \\
&\leq \left(\int_\Gamma k(y,y) d\mu \right)^2 \\
&\leq \left(\mu(\Gamma) \operatorname*{ess\,sup}_{y\in\Gamma} k(y,y) \right)^2 \\
&< +\infty
\end{aligned}$$

This proves that $DT : L^2(\Gamma, d\mu) \longrightarrow L^2(\Gamma, d\mu)$ is Hilbert-Schmidt and therefore compact. In addition it is self adjoint, and thus DT possesses eigenvectors and these form an orthonormal basis of $L^2(\Gamma, d\mu)$.

Appendix F. Bessel kernels

In the case of the Slepian prolates, the kernel is given by the inverse Fourier transform of the ball centered at the origin and of radius $\frac{c}{2}$. This kernel is radial, therefore if $x = re_n$ is along the last coordinate with $r \geq 0$, and if $c = 2$ and if we write $\xi = (\xi', \xi_n)$

$$\begin{aligned}
\int_{\|\xi\|<1} e^{2i\pi\langle\xi,x\rangle} d\xi &= \int_{\|\xi\|<1} e^{2i\pi r\xi_n} d\xi \\
&= \int_{\|\xi'\|<1} \int_{-\sqrt{1-\|\xi'\|^2}}^{\sqrt{1-\|\xi'\|^2}} e^{2i\pi r\xi_n} d\xi_n d\xi' \\
&= \int_{\|\xi'\|<1} \frac{\sin(2\pi r\sqrt{1-\|\xi'\|^2})}{\pi r} d\xi' \\
&= \frac{1}{\pi r}\int_{\mathbb{S}^{n-2}}\int_0^1 \sin(2\pi r\sqrt{1-\rho^2})\rho^{n-2} d\rho d\sigma \\
&= C\int_0^1 \sin(2\pi r\sqrt{1-\rho^2})\rho^{n-2} d\rho \\
&= C\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin(2\pi r\cos\phi)\sin^{n-2}\phi\cos\phi d\phi \\
&= C\left(\left[\sin(2\pi r\cos\phi)\frac{\sin^{n-1}(\phi)}{n-1}\right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\right. \\
&\quad \left. + \frac{2\pi r}{n-1}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(2\pi r\cos\phi)\sin^n\phi d\phi\right) \\
&= C'\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(2\pi r\cos\phi)\sin^n\phi d\phi
\end{aligned}$$

Now Poisson's Bessel function formula states that if $\Re(\nu) > -\frac{1}{2}$,

$$J_\nu(z) = \frac{1}{\sqrt{\pi}\Gamma(\nu+1)}\left(\frac{z}{2}\right)^\nu \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(z\cos\phi)\sin^{2\nu}\phi d\phi$$

This proves that

$$\int_{\|\xi\|<1} e^{2i\pi\langle\xi,x\rangle} d\xi = C''\frac{J_{\frac{n}{2}}(2\pi r)}{r^{\frac{n}{2}}}$$

Using that

$$\lim_{z\to 0}\frac{J_\nu(z)}{z^\nu} = \frac{\pi^\nu}{2^\nu\Gamma(\nu+1)}$$

and equating both hands of above equation to the volume of the unit ball yields $C'' = 1$. We find (after rescaling by $\frac{c}{2}$) that the kernel is

$$k(x,y) = \left(\frac{c}{2}\right)^{\frac{n}{2}}\frac{J_{\frac{n}{2}}(\pi cr)}{r^{\frac{n}{2}}}$$

where $r = \|x - y\|$.

In the case where n is odd, this can be further expressed in terms of derivatives of the sinc function: the spherical Bessel function of the first kind is

$$j_{\frac{n-1}{2}}(t) = \sqrt{\frac{\pi}{2t}} J_{\frac{n}{2}}(t)$$

and it can be proved that

$$j_p(t) = (-1)^p t^p \left(\frac{1}{t}\frac{d}{dt}\right)^p \frac{\sin t}{t}$$

therefore,

$$k(x,y) = \left(\frac{c}{2}\right)^{\frac{n}{2}} \sqrt{2c}(-1)^{\frac{n-1}{2}} \left(\frac{1}{r}\frac{d}{dr}\right)^{\frac{n-1}{2}} \mathrm{sinc}(rc)$$

References

[Aronszajn] N. Aronszajn, "Theory of Reproducing Kernels", *Transactions of the American Mathematical Society'*, Volume 68, No 3, May 1950.

[BellLabs1] D. Slepian and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty I", *the Bell System Technical Journal*, volume 40, pp 43-64, January 1961.

[BellLabs2] H. J. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty II", *the Bell System Technical Journal*, volume 40, pp 65-84, January 1961.

[BellLabs3] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty IV: Extensions to Many Dimensions; Generalized Prolate Spheroidal Wave Functions", *the Bell System Technical Journal*, volume 43, pp 3009-3058, November 1964.

[BellLabs5] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty V: The discrete case", *the Bell System Technical Journal*, volume 57, pp 1371-1430, may-jun 1978.

[Bochner] S. Bochner, "Hilbert Spaces and Positive Definite Functions", *The Annals of Mathematics*, 2nd series, Volume 42, No 3, pp 647-656, July 1941.

[Landau] H. J. Landau, "On Szegö's Eigenvalue Distribution Theorem and Non-Hermitian Kernels", *Journal d'Analyse Mathématique*, volume 28, 1975.

[Schoenberg] I. J. Schoenberg, "Metric spaces and Positive Definite Functions", *Transactions of the American Mathematical Society'*, Volume 44, No 3, pp 522-536, November 1938.

[YWeiss] Y. Weiss, "Segementation using Eigenvectors: a Unifying View", *International Conference in Computer Vision*, Volume 2, September 1999.

R. R. Coifman, Department of Applied Mathematics, Yale University, New Haven, CT 06511 USA

E-mail address: `coifman-ronald@yale.edu`

Stéphane Lafon, Department of Applied Mathematics, Yale University, New Haven, CT 06511 USA

E-mail address: `stephane.lafon@yale.edu`