Geometric Harmonics

Ronald R. Coifman, Stéphane Lafon

Applied Mathematics Department, Yale University, New Haven CT 06510, USA

Abstract

We describe a simple scheme, based on the Nyström method, for extending empirical functions f defined on a set X to a larger set $\overline{X}$. The extension process that we describe involves the construction of a specific family of functions that we term *geometric harmonics*. These functions constitute a generalization of the prolate spheroidal wave functions of Slepian in the sense that they are optimally concentrated on X. We study the case when X is a submanifold of $\mathbb{R}^n$ in greater detail. In this situation, any empirical function f on X can be characterized by its decomposition over the intrinsic Fourier modes, *i.e.*, the eigenfunctions of the Laplace-Beltrami operator, and we show that this intrinsic frequency spectrum specifies the largest domain of extension of f to the entire space $\mathbb{R}^n$.

Key words: Nyström method, intrinsic and extrinsic geometries, subsampling, prolate functions

1 Introduction

In applications where a large amount of data is involved, the only way to perform certain tasks like clustering, regression and classification is to subsample the data set $\overline{X}$ in order to reduce the size of the problem, process the new set X, and then extend the result to the original data $\overline{X}$. If in addition, the data are embedded in a high-dimensional space, then kernel methods like those based on radial functions constitute an interesting alternative to grid-based methods (which are unusable in high dimension).
In this paper, we describe a scheme for extending empirical functions defined on X to $\overline{X}$. This technique is inspired from the Nyström method, widely used in partial differential solvers, and recently employed in machine learning and

Email addresses: coifman@math.yale.edu (Ronald R. Coifman), stephane.lafon@yale.edu (Stéphane Lafon).

in spectral graph theory as a way to subsample large data sets (3; 6; 10). For instance, in order to perform an efficient (fast) image segmentation, it was noticed that one could subsample the pixels of an image, do the segmentation in this subgrid, and then extend the results to the entire image. In this article, we show how the Nyström method naturally leads to the construction of a particular set of functions that we term *geometric harmonics*, and we show that these functions, which are solutions of an eigenproblem and generalize the prolate spheroidal wave functions (8; 9), are optimally concentrated on X. In the case when X is a submanifold of $\mathbb{R}^n$, any function f can decomposed as a sum of intrinsic Fourier modes, namely, the eigenfunctions of the Laplace-Beltrami operator, and we show that the geometric harmonics relate the intrinsic oscillations of functions on X to that of their extensions to $\overline{X}$, and therefore build bridges between the intrinsic and extrinsic Fourier analyses. More precisely, we show that the restriction and extension operations preserve the bandwidth of signals, and that, following some version of the Heisenberg principle, a function f with intrinsic bandlimit ν on X can be extended as a function with the same extrinsic bandlimit on $\mathbb{R}^n$, and is numerically supported in a tubular neighborhood of radius $\mathcal{O}(\frac{1}{\nu})$ around X. This allows us to define a multiscale extension scheme for empirical functions in which each function is decomposed as a superposition of atoms with specific time-frequency localization.

The paper is organized as follows. In section 2, we introduce the notation, explain the construction of the geometric harmonics and their properties. We also describe the associated extension algorithm of empirical functions. In section 3, we give examples of geometric harmonics, and in particular we show that they generalize the prolate spheroidal wave functions of Slepian *et al.* Last, in section 4, we focus on the case when X is a submanifold of $\mathbb{R}^n$.

2 Geometric harmonics

2.1 Construction

Let X and $\overline{X}$ be two sets such that $X \subset \overline{X}$. Let μ be a finite measure on X,*i.e.*, $\mu(X) < +\infty$. Our goal is to be able to extend a function f defined on X to the set $\overline{X}$ and for that to be done, we construct a set of functions, the geometric harmonics, that allow us to perform this extension. Notation: functions defined on $\overline{X}$ will be denoted using capital letters, whereas functions on X will be represented using lower case letters. In this section, arbitrary points in $\overline{X}$ will be denote with a bar, for instance $\overline{x}$.

Our main ingredient is a "kernel" $k : \overline{X} \times \overline{X} \to \mathbb{R}$ satisfying

- k is symmetric: $k(\overline{x}, \overline{y}) = k(\overline{y}, \overline{x})$
- k is positive semi-definite, *i.e.*, for any $m \geq 1$ and any choice of real numbers $\alpha_1, ..., \alpha_m$, and of points $\overline{x}_1, ..., \overline{x}_m$, we have

  $$\sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j k(\overline{x}_i, \overline{x}_j) \geq 0 \,.$$

  This property is not necessary to be able to define geometric harmonics and extensions, however, it allows to interpret the geometric harmonics as maximizing some concentration measure over X.
- k is bounded on $\overline{X} \times \overline{X}$ by a number $M > 0$. Although this assumption can be weakened, it is very convenient for the simplicity of the exposition.

Since k is positive semi-definite, there exists a unique reproducing kernel Hilbert space $\mathcal{H}$ of functions defined on $\overline{X}$ for which k is the reproducing kernel (see appendix A). Let $\langle \cdot, \cdot \rangle_{\overline{X}}$ and $\| \cdot \|_{\overline{X}}$ be denote the inner product and norm of this space. In contrast, we use $\langle \cdot, \cdot \rangle_X$ and $\| \cdot \|_X$ to represent the inner product and norm in $L^2(X, d\mu)$.

The kernel k can be restricted to X, and employed to define a linear operator $\mathbf{K} : L^2(X, d\mu) \to \mathcal{H}$:

$$\mathbf{K}f(\overline{x}) = \int_X k(\overline{x}, x) f(x) d\mu(x) \text{ where } \overline{x} \in \overline{X} \,.$$

We have the following lemma:

Lemma 1 *The adjoint $\mathbf{K}^* : \mathcal{H} \to L^2(X, d\mu)$ is the restriction operator on X: for all $F \in \mathcal{H}$ and $x \in X$,*

$$\mathbf{K}^* F(x) = F(x) \,.$$

Moreover, $\mathbf{K}^*\mathbf{K} : L^2(X, d\mu) \to L^2(X, d\mu)$ is compact.

PROOF. First, in can be checked that the adjoint is given by $\mathbf{K}^* F(\overline{x}) = \langle k(\overline{x}, \cdot), F \rangle_{\overline{X}}$. This fact, together with the reproducing kernel identity (A.1) implies the first assertion. We now show that $\mathbf{K}^*\mathbf{K}$ is Hilbert-Schmidt, which implies compactness:

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

Indeed, we have

$$\int_X \int_X |k(x,y)|^2 d\mu(y) d\mu(x) \leq M^2 \mu(X)^2 < +\infty \,. \qquad \square$$

Since this operator is self-adjoint and compact on $L^2(X, d\mu)$, it has a discrete sequence of eigenvalues $\{\lambda_j\}$ (in non-increasing order) and eigenvectors $\{\psi_j\}$ defined on X: for $d\mu$-almost all $x \in X$,

$$\lambda_j \psi_j(x) = \int_X k(x,y)\psi_j(y) d\mu(y) \,.$$

Note that, because the operator is positive semi-definite, for all j, $\lambda_j \geq 0$. Also, by the result of this lemma, ψ_j is obtained by diagonalizing the kernel $k(x,y)$ with x and y restricted to X.
The eigenfunction ψ_j is defined on X, but provided $\lambda_j \neq 0$, it can be extended to the set $\overline{X}$ via a technique known as the Nyström method (see (7)). This technique consists in observing that $k(x,y)$ is defined for $x \in \overline{X}$, and therefore so is the right-hand side of the above equation, and that it can be used for the definition ψ_j outside X.

Definition 2 *When $\lambda_j \neq 0$, the eigenfunction ψ_j can be extended to $\overline{x} \in \overline{X}$ by*

$$\Psi_j(\overline{x}) = \frac{1}{\lambda_j} \int_X k(\overline{x}, y)\psi_j(y) d\mu(y) \,.$$

This extension is called Geometric Harmonic.

The term "geometric harmonics" is inspired from the fact that ψ_j is extended as an average of its values on the set X, and thus can be thought of as verifying a certain form of mean value theorem.
Numerically, this extension procedure is extremely ill-conditioned as one divides by the eigenvalues of a compact operator ($\lambda_j \to 0$ as $j \to +\infty$). Consequently, we introduce the following notations:

Definition 3 *For any $\delta > 0$, let*

$$S_\delta = \{j \text{ such that } \lambda_j \geq \delta \lambda_0\} \,,$$

and define the following finite-dimensional vector spaces

$$L^2_\delta = Span\{\psi_j, j \in S_\delta\} \,,$$

and

$$\mathcal{H}_\delta = Span\{\Psi_j, j \in S_\delta\}\,,$$

The extension procedure from L^2_δ to $\mathcal{H}_\delta$ has a condition number equal to $\frac{1}{\delta}$. We summarize the algebraic relation between $\mathbf{K}, \mathbf{K}^*, \Psi_j$ and ψ_j:

$\mathbf{K}\psi_j = \Psi_j$ (Extension),

$\mathbf{K}^*\Psi_j = \lambda_j\psi_j$ (Restriction).

We conclude this section with two remarks. First, an important class of positive kernels is generated by covariance kernels, *i.e.*, by kernels of the form

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

where $\{e_\xi\}_{\xi\in I}$ is a family of functions defined on $\overline{X}$ and $p(\xi) \geq 0$. Each function e_ξ, restricted to X, is interpreted as a vector whose coordinates are indexed by $x \in X$, and the kernel k represents the covariance of the cloud of points generated by the mass distribution $p(\xi)d\xi$. Finding the eigenfunctions and eigenvalues associated with this kernel is equivalent to computing the axes and moments of inertia of this cloud of points, which is also referred to Principal Component Analysis.
The other remark concerns a variational interpretation of k. Let $\mathbf{B}$ represent the orthogonal projector onto $\mathcal{H}$, defined by $\mathbf{B}F(x) = \langle F, k(x,\cdot)\rangle_{\mathcal{H}}$, and let $\mathbf{D}$ be the restriction operator to X defined by $\mathbf{D}F(\overline{x}) = F(\overline{x})$ if $\overline{x} \in X$, and $\mathbf{D}F(\overline{x}) = 0$ otherwise. Then it can be checked that $\mathbf{K} = \mathbf{BD}$ and if $x \in X$,

$$\mathbf{K}^*\mathbf{K}f(x) = \mathbf{DBD}f(x)\,.$$

This decomposition of $\mathbf{K}^*\mathbf{K}$ as a product of orthogonal projection leads to a variational interpretation of the geometric harmonics that motivated Slepian *et al* in introducing the prolate spheroidal wave functions (8; 9).

2.2 Properties

The geometric harmonics feature two interesting properties: they are orthogonal on X *and* on $\overline{X}$, and among all functions of $\mathcal{H}$, they have maximum concentration on X.

Proposition 4 *The system $\{\Psi_j\}_{j\in S_\delta}$ forms an orthogonal basis of $\mathcal{H}_\delta$, and*

their restrictions $\{\psi_j\}_{j\in S_\delta}$ *to* X *forms an orthogonal basis of* L^2_δ.

PROOF. By definition, the ψ_j's are obtained as the eigenfunctions of a self-adjoint operator, therefore they are orthogonal on X. In addition,

$$\begin{aligned}\langle \Psi_i, \Psi_j\rangle_{\overline{X}} &= \frac{1}{\lambda_j}\langle \Psi_i, \mathbf{K}\mathbf{K}^*\Psi_j\rangle_{\overline{X}} \\ &= \frac{1}{\lambda_j}\langle \mathbf{K}^*\Psi_i, \mathbf{K}^*\Psi_j\rangle_{\overline{X}} \\ &= \frac{1}{\lambda_j}\langle \psi_i, \psi_j\rangle_X\end{aligned}$$

□

Definition 5 *For a function* $F \in \mathcal{H}$, *with restriction* $f \in L^2(X, d\mu)$, *we define the concentration of* F *over* X *to be the Rayleigh quotient*

$$c_X(F) = \frac{\|f\|_X}{\|F\|_{\overline{X}}}\,.$$

where $f = \mathbf{K}^*F$ *is the restriction of* F *to* X.

The geometric harmonics are also the function of $\mathcal{H}$ that have maximum concentration on the set X:

Proposition 6 *The function* Ψ_j *is a solution to the problem*

$$\max_{F\in\mathcal{H}} c_X(F)\,,$$

under the constraint that $F \perp \{\Psi_0, \Psi_1, ..., \Psi_{j-1}\}$. *In particular,* Ψ_0 *is the element of* $\mathcal{H}$ *that is the most concentrated on* X.

PROOF. By homogeneity of the ratio, we can restrict our attention to all $F \in \mathcal{H}$ with norm 1. Thus we need to maximize

$$\langle f, f, \rangle_X = \langle \mathbf{K}^*F, \mathbf{K}^*F\rangle_X = \langle F, \mathbf{K}\mathbf{K}^*F\rangle_{\overline{X}}\,,$$

under the constraints $\langle F, F\rangle_{\overline{X}} = 0$, $\langle F, \Psi_0\rangle_{\overline{X}} = 0$,...,$\langle F, \Psi_{j-1}\rangle_{\overline{X}} = 0$. Using the Lagrange multipliers technique, we conclude that there exist numbers λ and $\alpha_0, ..., \alpha_{j-1}$ such that

$$\mathbf{K}\mathbf{K}^*F = \lambda F + \alpha_0\Psi_0 + ... + \alpha_{j-1}\Psi_{j-1}\,.$$

Taking the inner product with Ψ_i for $i = 0, ..., j-1$, and invoking the constraints and the orthogonality of the geometric harmonics (see previous proposition), we obtain that

$$\alpha_i \|\Psi_i\|_{\overline{X}}^2 = \langle \mathbf{K}\mathbf{K}^* F, \Psi_i \rangle_{\overline{X}} = \langle F, \mathbf{K}\mathbf{K}^* \Psi_i \rangle_{\overline{X}} = \lambda_i \langle F, \Psi_i \rangle_{\overline{X}} = 0 \,.$$

As a consequence, $\alpha_i = 0$, and F is a geometric harmonic associated with the eigenvalue λ, and the functional to be minimized now takes the form

$$\langle f, f \rangle_X = \lambda \,.$$

The maximum is therefore achieved for $F = \Psi_j$. $\square$

2.3 Extension Algorithm

We now describe the natural extension algorithm associated with the geometric harmonics. We assume that ψ_j is normalized so that it has norm 1 on X.

Algorithm 1 (Extension Scheme) *Given a function $f \in L^2(X, d\mu)$,*

- *project f onto the space L^2_δ spanned by the orthonormal system $\{\psi_j\}_{j \in S_\delta}$:*

$$f \mapsto \mathbf{P}_\delta f = \sum_{j \in S_\delta} \langle f, \psi_j \rangle_X \psi_j \,,$$

- *use the extension Ψ_j of ψ_j to extend $\mathbf{P}_\delta f$ on $\overline{X}$ as*

$$\mathbf{E}f(\overline{x}) = \sum_{j \in S_\delta} \langle f, \psi_j \rangle_X \Psi_j(\overline{x}) \,,$$

with $\overline{x} \in \overline{X}$.

This algorithm computes a truncated pseudo-inverse for K^*, and is consistent in the sense that the restriction of $\mathbf{E}f$ to X is, again, extended as $\mathbf{E}f$ by the algorithm.

We now make two comments on this algorithm. First, this technique does not provide an extension for f but rather for a filtered version of it, namely, its orthogonal projection $\mathbf{P}_\delta f$ onto L^2_δ. This space is precisely that of functions that can be numerically extended to $\overline{X}$, since the condition number of the operator $\mathbf{E}$ is $\frac{1}{\delta}$. Moreover, a general empirical function defined on X can be

extended if the residual $\|f - \mathbf{P}_\delta f\|_X$ is smaller than a prescribed error, which leads to the definition:

Definition 7 *A function f defined on X is said to be (η, δ)-extendable if*

$$\|\mathbf{P}_\delta f\|_X = \sum_{j \in S_\delta} |\langle f, \psi_j \rangle_X|^2 \geq (1 - \eta)\|f\|_X^2 \,.$$

The obvious consequence is that not all functions f can be extended at a given precision, and this fact can be used to relate the oscillations of f on X to a notion of extrinsic bandlimit (see section 4.

The second point we wish to discuss concerns the interpretation of the extension. As we shall see in the next section, for a given f defined on X, there generally exist infinitely many possible extensions $F \in \mathcal{H}$ as functions in $\mathcal{H}$ might not be determined by their values on X. However, according to proposition 6, $\mathbf{E}f$ is the extension that has maximal concentration over X. From the point of view of statistics, where extension means regression, this is related to the predictive ability of Nyström extensions. For instance, it is known that if a random variable $(U, V) \in \mathbb{R}^{m+n}$ has the following covariance matrix

$$C(U, V) = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$$

then the covariance of V given U is

$$C(V|U) = \begin{pmatrix} A & B \\ B^T & C - B^T A^{-1} B \end{pmatrix}$$

while the Nystöm extension method (taking X to be the set of first m indices, and $\overline{X}$ being the set of $m + n$ indices) implicitly approximates $C(U, V)$ as

$$C(U, V) \simeq \begin{pmatrix} A & B \\ B^T & B^T A^{-1} B \end{pmatrix}$$

Therefore, the norm of the Schur complement $\|C - B^T A^{-1} B\|$ is a measure of the average randomness of V that is left when one knows U. If the Nyström approximation is accurate, then the predictive power of U over V is large.

3 Examples of geometric harmonics

In this section we give some examples of geometric harmonics.

3.1 The prolate spheroidal wave functions - bandlimited extension

In (8), Slepian *et al* introduce the prolate spheroidal wave functions as the solution of finding functions optimally concentrated in time and frequency. By construction, the prolates are bandlimited functions of unit energy that have maximum energy within an interval of the time domain. They also generalize their result to higher dimension (9) by defining $\mathcal{H}^B$ to be the space of functions of $L^2(\mathbb{R}^n$ whose Fourier is supported in the ball centered at the origin and of radius $\frac{N}{2}$. This space is a reproducing kernel Hilbert space with kernel (see appendix B)

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

where J_ν is the Bessel function of the first kind and of order ν. Such a kernel is referred to as a Bessel kernel. The prolate spheroidal wave functions are defined as the eigenfunctions of the integral operator of kernel k_B restricted to a certain space domain $X \subset \overline{X} = \mathbb{R}^n$. In the prolate setting, X is a set of positive measure (and the kernel is non-integrable), and functions of $\mathcal{H}^B$ are determined by their values on X. On the contrary, we are mainly interested in the case where X is singular in $\mathbb{R}^n$. In this situation, there are infinitely many bandlimited extensions for a function f defined on X, as two such extensions differ by a (non-zero) bandlimited function that vanishes on X. However, among all these extensions, the operator $\mathbf{E}$ produces an extension with maximum concentration on X, or equivalently, that has minimal energy outside X.
From now on, we will refer to $\mathbf{E}_B$ as the Bessel kernel with bandlimit $B > 0$; this operator computes the bandlimited extension of band B that is maximally concentrated on X. On figure 1, we show examples of bandlimited extensions (at a fixed band) of the functions $\cos(j\theta)$ for $j = 1, 2, 4, 8$ from the unit circle to the plane. The Bessel kernel tends to generate a large amount of oscillations outside the circle, whereas inside the circle the oscillation seem to cancel out.

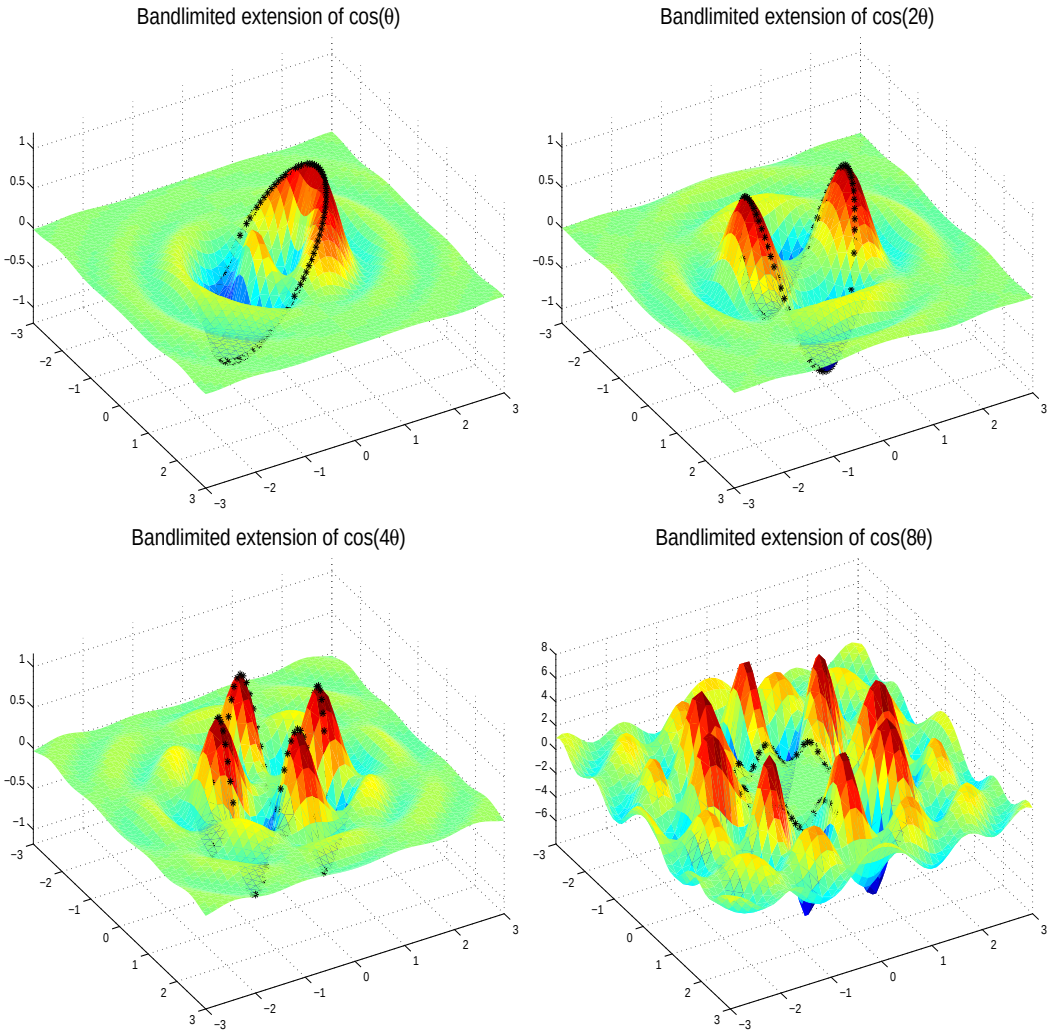

Fig. 1. Bandlimited extensions of the functions $\cos(j\theta)$ for $j = 1, 2, 4, 8$, from the unit circle to the plane.

3.2 Harmonic extension

Another example comes from potential theory. Consider the single-layer Newtonian potential in $\overline{X} = \mathbb{R}^n$,

$$k(x,y) = \begin{cases} -\log(\|x-y\|) & \text{if } n = 2\,, \\ \frac{1}{\|x-y\|^{n-2}} & \text{if } n \geq 3\,. \end{cases}$$

By definition, this kernel is the Green's function for the Laplace operator on $\mathbb{R}^n$, and as a consequence, the potential is positive semi-definite. Assuming that $n = 3$, and that X is a Lipschitz surface in $\mathbb{R}^3$, then $\mathcal{H}$ is the space of

potentials

$$F(x) = \int_X \frac{d\rho(y)}{\|x-y\|}\,,$$

where ρ is a signed measure supported on X, representing a distribution of charges. In this space, the inner product is the electrostatic energy of interaction between two distributions of charges:

$$\langle F_1, F_2 \rangle_{\overline{X}} = \int_X \int_X \frac{d\rho_1(y) d\rho_2(x)}{\|x-y\|}\,.$$

Note that all elements of $\mathcal{H}$ are harmonic and that the geometric harmonics are harmonic functions that minimize their electrostatic self-energy. The operator $\mathbf{E}$ allows to construct harmonic extensions for empirical functions defined on X.

An interesting feature of the harmonic extension is that, unlike the bandlimited extension, it does not involve a scale (or bandwidth) parameter. In fact, the size of the numerical support of the extension of f depends on the oscillations of f on X. Indeed, suppose that X is a closed curve in $\mathbb{R}^3$, and that f oscillates p times along X. Then since f is extended as a sum of elementary potentials generated by charges along X, it is roughly extended as the potential generated by a multipole of order p, and therefore will decay as $\frac{1}{r^p}$ where r is the distance from the curve X. This relation between the intrinsic frequencies of f and its domain of extension is illustrate on figure 2 and is further investigated in section 4.

3.3 Wavelet extension

Let $\{V_j\}_{j\in\mathbb{Z}}$ be a multiresolution analysis in $\mathbb{R}^n$ and let $\mathcal{H}$ be the space of scaling functions V_j at a fixed scale 2^{-j}. Let X be a subset of $\mathbb{R}^n$. One can define geometric harmonics corresponding to this space and use them to construct scaling functions adapted to the geometry of X and to compute extensions at a given scale.

We illustrate this idea in the simple case of the Haar multiresolution. Let Φ be the indicator function of the unit cube in $\mathbb{R}^n$, and let X be a finite length curve. Therefore, for $\mathcal{H} = V_j$, the reproducing kernel is given by

$$\sum_{m\in\mathbb{Z}^n} 2^{-nj}\Phi(2^j x - m)\Phi(2^j y - m)\,,$$

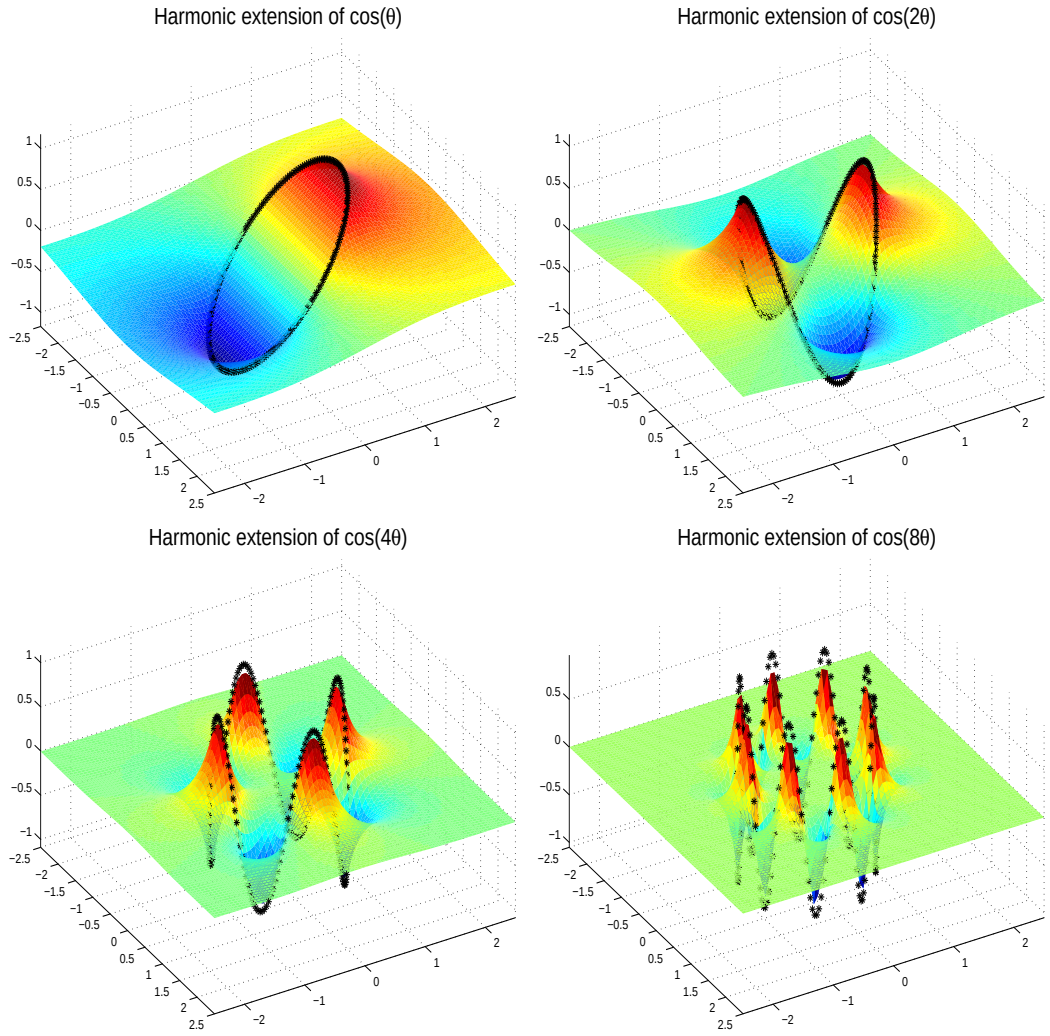

Fig. 2. Harmonic extensions of the functions $f(\theta) = \cos(j\theta)$ for $j = 1, 2, 4, 8$, from the unit circle to the plane. When the number of the oscillations increases, so does the order of the multipole extending f, and the extension decays faster.

and when one restricts it to the curve X,

$$k(x, y) = \sum_{m \in Q_X} 2^{-nj} \Phi(2^j x - m) \Phi(2^j y - m) \,,$$

where Q_X is the set of indices in $\mathbb{Z}^n$ associated with unit cubes intersecting X. This formula allows to conclude that the geometric harmonics are the indicator of the dyadic cubes at scale 2^{-j} that intersect the curve X, and it can be checked that the eigenvalues are given by the quantities $2^{-nj}|Q \cap X|$ where $|Q \cap X|$ is the length of the piece of X that intersect a given dyadic cube Q at scale 2^{-j}. As an illustration, on figure 3 we show the extension of the function $f(\theta) = \theta$ from the circle to the plane, at different scales.

The Haar case was simple as these scaling functions, when restricted to any

curve, remain orthogonal. For more general scaling functions, the geometric harmonics provide scaling functions adapted to the geometry of X. Note that the same kind of construction can be done with wavelet spaces W_j.

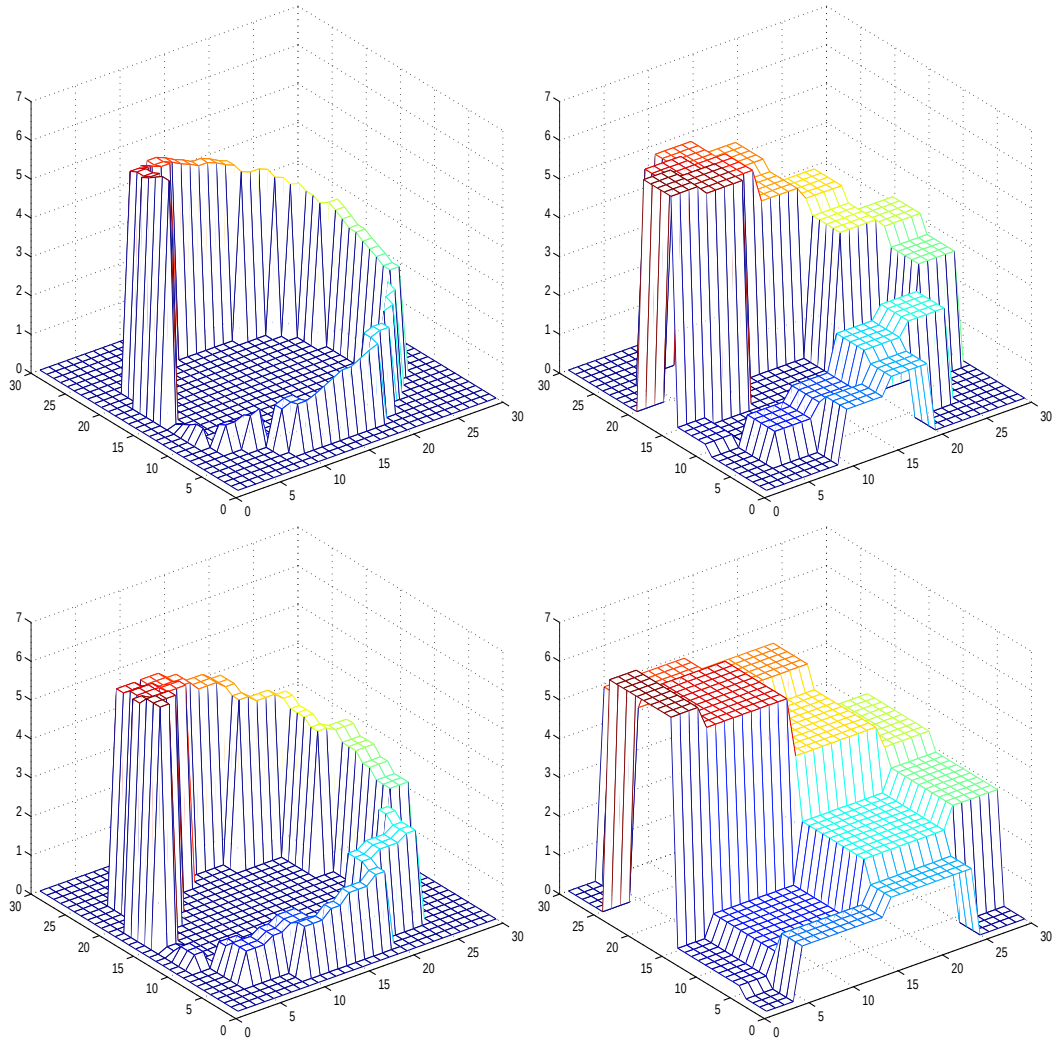

Fig. 3. Extension of the function $f(\theta) = \theta$ using the Haar scaling function at different scales.

4 Multiscale extension

We now restrict our attention to the case when $\overline{X} = \mathbb{R}^n$ and X is smooth compact submanifold of dimension d. To simplify the exposition, we assume that the measure $d\mu$ is now the Riemannian measure dx on X. In this case, two different Fourier (or Paley-Littlewood) analyses can be performed on a function f defined on X. The first one is purely intrinsic and is obtained by using the eigenfunctions of the Laplace-Beltrami operator Δ on X. We recall that these eigenfunctions are the analogue of the Fourier basis to arbitrary sub-

manifolds. The spectrum of Δ is discrete (as X is compact) and corresponds to pure frequency modes. The other frequency analysis is that of the classical Fourier transform in $\mathbb{R}^n$ and that can be applied to the various extensions of f.

In this section we investigate the relation between these two analyses by exploring the action of the restriction and extension operators. In particular, we ask the question of whether the intrinsic diffusion of heat is equivalent to the extrinsic diffusion. We show that this is true, provided the embedding of the manifold X into $\mathbb{R}^n$ is not too "complicated". We also develop a multiscale extension scheme for functions defined on X.

4.1 Restriction operator

Since the set X is a smooth submanifold, the restriction of a function F to X is a well-behaved operation, in the sense that if F is smooth, then so is its restriction f.

We can give a precise meaning of this statement using a space characterization of smoothness: suppose that F is differentiable with a bounded derivative, then f is obviously differentiable as a map from X into $\mathbb{R}$, and its (intrinsic) gradient at a point $x \in X$ is nothing else but the orthogonal projection of gradient of F onto the tangent plane at that same point x.

The same idea can be characterized by a frequency argument. Consider plane waves in $\mathbb{R}^n$:

$$F_\xi(x) = e^{2i\pi\langle \xi, x\rangle}$$

Let Δ be the Laplace-Beltrami operator on X that we can assume to be a curve for the sake of simplicity. Let $\{\nu_j^2\}$ and $\{\phi_j\}$ be its eigenfunctions (in non-decreasing order) and eigenvalues:

$$\Delta\phi_j = \nu_j^2\phi_j\,.$$

Definition 8 (Intrinsic bandwidth) *A function $f \in L^2(X, dx)$ defined on X is said to be* intrinsically bandlimited *if it can be written as*

$$f = \sum_{j=0}^{m} c_j\phi_j\,,$$

for some $m \geq 0$, with $c_m \neq 0$. The largest pure frequency ν_m^2 in this decomposition is called intrinsic bandwidth *of f.*

To relate the extrinsic bandwidth ξ to the intrinsic one, we have the following result:

Proposition 9 *Suppose that the curvature of X is bounded by a number $M > 0$.*
Then if $\|\xi\| > \frac{M}{\pi}$, the intrinsic spectrum of f_ξ decays exponentially according to

$$|\langle \phi_j, f_\xi \rangle| \leq \sqrt{\mu(X)} \left(\frac{4\pi^2 \|\xi\|^2}{\nu_j^2} \right)^m$$

for all $m \geq 0$. As a conclusion, a function with only low extrinsic frequencies is also a function with low intrinsic frequencies when restricted to X, with approximately the same band.

PROOF. Locally, on the curve around $x \in X$, the function F_ξ has the form

$$f_\xi(y) = F_\xi(y) = \exp(2i\pi(\langle \xi, x \rangle + u\xi_T + a(x)u^2\xi_N)$$

where u is the local coordinate in the tangent plane of the point y, $a(x)$ is the scalar curvature at x, and ξ_T and ξ_N are the tangent and normal projections of ξ in the osculatory plane at x (see configuration on figure 4). A Taylor

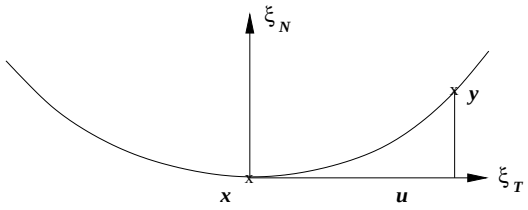

Fig. 4. Geometric configuration.

expansion yields:

$$f_\xi(y) = e^{2i\pi\langle \xi, x \rangle} \left(1 + 2i\pi\xi_T u + 2i\pi(a(x)\xi_N + i\pi\xi_T^2)u^2 + ... \right) .$$

Now, since for any function f defined on X, we have the Taylor expansion

$$f(y) = f(x) + u\frac{\mathrm{d}f}{\mathrm{d}s}(x) + \frac{1}{2}u^2\frac{\mathrm{d}^2 f}{\mathrm{d}s^2}(x) + ... ,$$

and since $\Delta = \frac{\mathrm{d}^2}{\mathrm{d}s^2}$, we identify

$$\Delta f_\xi(x) = 4i\pi(a(x)\xi_N + i\pi\xi_T^2)f_\xi(x) .$$

Therefore, if the curvature is bounded by $M > 0$, we have the trivial estimate for $\|\xi\| > \frac{M}{\pi}$,

$$\|\Delta f_\xi\|_X \leq 4\pi^2\|\xi\|^2\|f_\xi\|_X .$$

In fact it is easily seen that for all $m \geq 0$,

$$\|\Delta^m f_\xi\|_X \leq \left(4\pi^2\|\xi\|^2\right)^m \|f_\xi\|_X \,.$$

Since

$$\Delta^m f_\xi(x) = \sum_{j\geq 0} \nu_j^{2m} \langle \phi_j, f\rangle \phi_j(x)\,,$$

and

$$\|f_\xi\|_X^2 \leq \mu(X)\,,$$

we must have, by Parseval,

$$\sum_{j\geq 0} \nu_j^{4m} |\langle \phi_j, f_\xi\rangle|^2 \leq \mu(X) \left(4\pi^2\|\xi\|^2\right)^{2m} \,.$$

In particular,

$$|\langle \phi_j, f_\xi\rangle| \leq \sqrt{\mu(X)} \left(\frac{4\pi^2\|\xi\|^2}{\nu_j^2}\right)^m$$

for all $m \geq 0$. Therefore the coefficients of expansions in the eigenfunctions of the Laplace-Beltrami operator are negligible, except for finitely many, namely those for which the eigenvalue ν_j^2 is less than $4\pi^2\|\xi\|^2$. $\square$

4.2 Extension Operator

The extension algorithm described in Section 2.3 is a two-step procedure

- the function f on the data is first pre-filtered by a projection on the geometric harmonics that numerically admit an extension,
- then the extension is computed.

However, we know that not all functions f defined on X are (η, δ)-extendable. This means that some functions cannot be extended to some bandwidth, but for a given (η, δ) and any function f, if we choose the band B sufficiently large, f will be (η, δ)-extendable as a B-bandlimited function. This leads to the natural definition for the notion of extrinsic bandwidth as the infimum

over all such $B > 0$. However, we prefer the following alternate definition which is easier to deal with:

Definition 10 (Extrinsic bandwidth) *For a fixed value of $\varepsilon > 0$, the extrinsic bandwidth of a function f defined on X is the infimum of all $B > 0$ such that there exists a function F defined on $\mathbb{R}^n$ satisfying:*

- *F is an extension of f, i.e.,*

$$F(x) = f(x) \text{ for all } x \in X\,,$$

- *F can be approximated to relative precision ε by a B-bandlimited function G on $\mathbb{R}^n$:*

$$\frac{\left(\int_{\mathbb{R}^n} |F_j(x) - B_j(x)|^2\, dx\right)^{\frac{1}{2}}}{\left(\int_{\mathbb{R}^n} |F_j(x)|^2\, dx\right)^{\frac{1}{2}}} \leq \varepsilon\,.$$

For the study of the restriction operator, we looked at the restriction of the Fourier modes in $\mathbb{R}^n$, namely the plane waves. For the extension problem, it is natural to extend the Fourier modes on X, i.e. the set of eigenfunctions of the Laplace-Beltrami operator on X. In order to relate the intrinsic frequencies that these functions represent to the extrinsic spectrum (provided by the Fourier analysis in $\mathbb{R}^n$), it can be instructive to compute the extrinsic bandwidth of ϕ_j, or equivalently how far ϕ_j can be extended away from the set.

A simple example displayed on figure 5 shows that if the embedding of the set X in $\mathbb{R}^n$ is complicated, then the extrinsic bandwidth can be very different from the intrinsic bandwidth. The curve X was chosen to have low frequencies on the left part, and steep variations on the right part. We considered the following intrinsic wave packets:

$$w_x^t(y) = w_{x,t}(y)\phi_j(y)$$

where $w_{x,t}$ is a window function centered at x and of width $\sqrt{t}$. For instance,

$$w_{x,t}(y) = \exp\left(-d^2(x,y)/t\right)$$

is a Gaussian window along the curve, where $d(x,y)$ is the geodesic distance between x and y. In some sense, $w_x^{j,t}$ defines an intrinsic local cosine waveform: x represents the time location parameter, j is the intrinsic frequency location, and $\sqrt{t}$ is the time-width. On figure 5 we have plotted the domains of the plane where the function $w_x^{j,t}$ can be locally extended, for different values of x, and a fixed value of j and t.

Clearly, this elementary example shows that the extrinsic bandwidth can be much larger than the intrinsic one, especially at locations where the curve

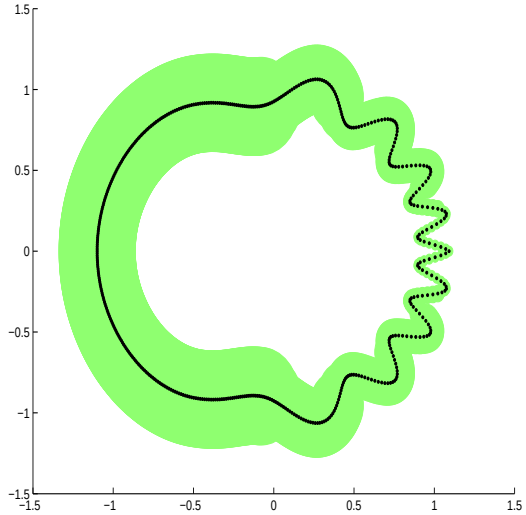

Fig. 5. Domain of extension for $w_x^{j,t}$ can be locally extended, for different values of x, and a fixed value of $j = 10$ and t.

has wild oscillations. In the following proposition, we show that a function f with intrinsic bandlimit B on X admits an approximate extension that is a bandlimited function with band CB where C is some universal constant that depends on the geometry of X. In order to adopt a broader point of view, instead of considering the eigenfunctions of the Laplace-Beltrami operator Δ, we will study the extension of eigenfunctions of the following elliptic operator:

$$\mathbb{A} = \Delta + Q\,,$$

where Q is a bounded potential function. In (4), the authors show that this type of differential operator arises naturally as the small-scale limit of several families of kernel operators. We keep the same notations for the eigenfunctions and eigenvalues, namely,

$$\mathbb{A}\phi_j = \nu_j^2 \phi_j\,.$$

Proposition 11 *Let $\varepsilon > 0$ be a preset accuracy. There exists a constant $C > 0$ such that for all $j \geq 0$, one can construct a function F_j defined on $\mathbb{R}^n$ satisfying:*

- *F_j is an extension of ϕ_j, i.e.,*

$$F_j(x) = \phi_j(x) \text{ for all } x \in X\,,$$

- *F_j can be approximated to relative precision ε by a bandlimited function B_j*

of band $C\nu_j$:

$$\frac{\left(\int_{\mathbb{R}^n} |F_j(x) - B_j(x)|^2 \, dx\right)^{\frac{1}{2}}}{\left(\int_{\mathbb{R}^n} |F_j(x)|^2 \, dx\right)^{\frac{1}{2}}} \leq \varepsilon \, .$$

In particular, this proposition asserts that the extrinsic bandwidth is less than or equal to the intrinsic bandwidth times a constant C that depends on the precision ε and on the geometry of X. In some way, it is a measure of the complexity of the embedding of X into $\mathbb{R}^n$. For instance, in the example of figure 5, C can be quite large because of the oscillations of the curve.

PROOF. For any $x \in \mathbb{R}^n$, let $x' \in X$ be such that

$$\|x - x'\| = \inf_{y \in X} \|x - y\| \, .$$

Define F_j by

$$F_j(x) = e^{-\nu_j^2 \|x - x'\|^2} \phi_j(x') \, .$$

The function F_j is an extension of ϕ_j to $\mathbb{R}^n$. To estimate the decay of its spectrum, we need to bound its gradient $\nabla^m F_j(x)$ (tensor of order m). The Leibnitz formula yields:

$$\|\nabla^m F_j(x)\| \leq \sum_{i=0}^{m} \binom{m}{i} \|\nabla^i (e^{-\nu_j^2 \|x - x'\|^2})\| . \|\nabla^{m-i}(\phi_j(x'))\| \, .$$

The triangle inequality in $L^2(\mathbb{R}^n)$ gives

$$\left(\int_{\mathbb{R}^n} \|\nabla^m F_j(x)\|^2 dx \right)^{\frac{1}{2}} \leq \sum_{i=0}^{m} \binom{m}{i} \left(\int_{\mathbb{R}^n} \|\nabla^i (e^{-\nu_j^2 \|x - x'\|^2})\|^2 . \|\nabla^{m-i}(\phi_j(x'))\|^2 dx \right)^{\frac{1}{2}} \, .$$

To evaluate each term of the right-hand side, we make use of the following lemma:

Lemma 12 *Let f_ν be a function on $\mathbb{R}^n$ of the form*

$$f_\nu(x) = g(\nu \|x - x'\|) h_\nu(x') \, ,$$

where g has an exponential decay. Again, let $M > 0$ be a bound on the curvature of X. Then, if $\nu > 4M$,

$$\int_{\mathbb{R}^n} |f(x)|dx \asymp \nu^{-(n-d)} \int_0^{+\infty} |g(r)|r^{n-d}dr \int_X |h_\nu(u)|du\,.$$

Because of the decay of g, up to exponentially small terms, this integral can be computed on the set Ω_ν of all points at distance less than or equal to a multiple of $\frac{1}{\nu}$. We can associate to any $x \in \mathbb{R}^n$ a pair (u,t) where $x' = u$ is the closest point to $x \in \Omega_\nu$ and $t = x - u$. Conversely, to any $u \in X$ and t normal to X at u, we can associate the point $x = u+t$. Let $J(u,t)$ denote the Jacobian of the change of variable $(u,t) \mapsto x$. The lemma follows from the fact that J is bounded from below and above for all $x \in \Omega_\nu$. Indeed, first, a variation dt of t entails the same variation of x. Second, a variation du in the tangent plane at u entails a variation of x of order $(1 + 2\alpha(u)|\|t\|)du$ in the tangent direction, where $|\alpha(u)| \leq M$, and of order $\|t\|du^2$ in the normal direction. To conclude, since $\|t\| < \frac{1}{\nu} < \frac{1}{4M}$, we obtain that $1 - 2\alpha(u)\|t\| > 1 - \frac{1}{2}$ and $1 + 2\alpha(u)\|t\| < 1 + \frac{1}{2}$. Finally,

$$\frac{1}{2} < J(u,t) < \frac{3}{2}\,.$$

Therefore, with the same constants,

$$\int_{\mathbb{R}^n} |f_\nu(x)|dx \asymp \int_X |h_\nu(u)|du \int_{\mathbb{R}^{n-d}} g(\nu\|t\|)dt\,,$$

which ends the proof of the lemma.

Going back to the proof of the proposition, the lemma implies that

$$\left(\int_{\mathbb{R}^n} \|\nabla^m F_j(x)\|^2 dx\right)^{\frac{1}{2}} \leq \sum_{i=0}^{m} \binom{m}{i} K_i \nu_j^{i-\frac{n-d}{2}} \left(\int_X \|\nabla^{m-i}\phi_j(u)\|du\right)^{\frac{1}{2}}, \quad (1)$$

where K_i is a constant of the order of magnitude of the L^2 norm of the i^{th} derivative of the univariate Gaussian at scale 1. What remains to be done is to bound the $L^2(X)$-norm of the derivatives of ϕ_j. To do so, we need the following result:

Lemma 13 *For all $s \geq 0$, there exists C'_s such that*

$$\|\phi_j\|_s = \left(\sum_{|\alpha| \leq s} \int_X |\partial^\alpha \phi_j(u)|^2 du \right)^{\frac{1}{2}} \leq C'_s \nu_j^i \,.$$

This lemma follows from the classical theory of elliptic operators, which says that since we can bound the norm of $\mathbf{\Delta}^k \phi_j$, we have a bound on all derivatives of order less than or equal to $2k$.

Let

$$C_Q = \sup_{u \in X} |Q(u)| \,.$$

For $i = 0$, the lemma is trivial, and for $i = 1$, it results from an integration by parts as

$$\begin{aligned} \|\nabla \phi_j\|_X^2 &= \langle \Delta \phi_j, \phi_j \rangle_X \quad \text{by the Stokes formula,} \\ &= \langle \mathbf{\Delta} \phi_j, \phi_j \rangle_X - \langle Q \phi_j, \phi_j \rangle_X \,, \\ &\leq \nu_j^2 + C_Q \,, \end{aligned}$$

and therefore

$$\|\phi_j\|_1^2 = \|\phi_j\|_X^2 + \|\nabla \phi_j\|_X^2 = 1 + \nu_j^2 + C_Q \leq C_1'^2 \nu_j^2 \,.$$

In (5), p 262, it is shown that if L is an elliptic operator of order k, then for all $i \geq k$ and all f defined on X, we have:

$$\|f\|_i \leq C'(\|Lf\|_{i-k} + \|f\|_{i-1}) \,. \tag{2}$$

We can now proceed by induction:

- for $i = 2s$ and $L = \mathbf{\Delta}^s$, identity (2) yields

$$\begin{aligned} \|\phi_j\|_{2s} &\leq C'(\|\mathbf{\Delta}^s \phi_j\|_X + \|\phi_j\|_{2s-1}) \\ &\leq C'(\nu_j^{2s} + C'_{2s-1} \nu_j^{2s-1}) \\ &\leq C'_{2s} \nu_j^{2s} \,, \end{aligned}$$

- for $i = 2s + 1$ and $L = \mathbf{\Delta}^s$, identity (2) becomes

$$\|\phi_j\|_{2s+1} \leq C'(\|\mathbf{\Delta}^s \phi_j\|_1 + \|\phi_j\|_{2s}) \,.$$

We have

$$\begin{aligned}
\|\blacktriangle^s\phi_j\|_1^2 &= \|\blacktriangle^s\phi_j\|_X^2 + \|\nabla\blacktriangle^s\phi_j\|_X^2 \quad \text{by definition,}\\
&= \nu_j^{4s} + \langle\Delta\blacktriangle^s\phi_j, \blacktriangle^s\phi_j\rangle_X \quad \text{by the Stokes formula,}\\
&= \nu_j^{4s} + \langle\blacktriangle^{s+1}\phi_j, \blacktriangle^s\phi_j\rangle_X - \langle Q\blacktriangle^s\phi_j, \blacktriangle^s\phi_j\rangle_X\,,\\
&\leq \nu_j^{4s} + \nu_j^{2(2s+1)} + C_Q\nu_j^{4s}\,,
\end{aligned}$$

and finally,

$$\|\phi_j\|_{2s+1} \leq C'\left(\sqrt{(1+C_Q)\nu_j^{4s} + \nu_j^{2(2s+1)}} + C'_{2s}\nu_j^{2s}\right) \leq C'_{2s+1}\nu_j^{2s+1}\,.$$

The lemma is now proven, and it allows us to finish the proof of the proposition as from equation (1), we can conclude that

$$\left(\int_{\mathbb{R}^n} \|\nabla^m F_j(x)\|^2 dx\right)^{\frac{1}{2}} \leq C_m \nu_j^{m-\frac{n-d}{2}}\,.$$

Now for a fixed value of m, define

$$\widehat{B}_j(\xi) = \begin{cases} \widehat{F}_j(\xi) & \text{if } \|\xi\| < C\nu_j,\\ 0 & \text{otherwise,}\end{cases}$$

then, by the Parseval identity, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} |F_j(x) - B_j(x)|^2\, dx &= \int_{\|\xi\|>C\nu_j} |\widehat{F}_j(\xi)|^2 d\xi\,,\\
&\leq \int_{\|\xi\|>C\nu_j} |\widehat{F}_j(\xi)|^2 \frac{\|\xi\|^{2m}}{(C\nu_j)^{2m}} d\xi\,,\\
&\leq \frac{1}{(C\nu_j)^{2m}} \int_{\mathbb{R}^n} \|\nabla^m F_j(x)\|^2 dx\,,\\
&\leq \frac{C_m^2}{C^{2m}} \nu_j^{-(n-d)}\,.
\end{aligned}$$

Form lemma 12, we have

$$\int_{\mathbb{R}^n} |F_j(x)|^2 dx \geq K^2 \nu_j^{-(n-d)}$$

for some $K > 0$, and we merely have to pick C so that

$$\frac{C_m^{\frac{1}{m}}}{KC} < \delta\,. \qquad \square$$

Recall that $\mathbf{E}_B$ is the bandlimited extension operator corresponding to band B. From now on, we set $B = C\nu_j$. The consequence of this proposition is that, because of its optimal property, the extension provided by $\mathbf{E}_B$ must have an energy on $\mathbb{R}^n$ that is less than or equal to that of the extension that we constructed in the proof above. This means that the numerical support of $\mathbf{E}_B\phi_j$ will be included in a tube of radius proportional to $\frac{1}{\nu_j}$ around X. Theoretically, it could be much thinner, but because of the Heisenberg principle, then the support cannot really be smaller:

Lemma 14 *The standard deviation of the extension $\mathbf{E}_B\phi_j$ along any normal direction to X is at least equal to $\frac{C'}{\nu_j}$ for some $C' > 0$ independent of ν_j.*

PROOF. Let f be the restriction of $\mathbf{E}_B\phi_j$ on a line that is normal to X. Then f is a univariate bandlimited function of band $C\sqrt{d}\nu_j$. Let

$$\mathrm{Var}(f) = \frac{\int_{\mathbb{R}}(x-\overline{x})^2|f(x)|^2dx}{\int_{\mathbb{R}}|f(x)|^2dx}$$

and

$$\mathrm{Var}(\widehat{f}) = \frac{\int_{\mathbb{R}}(\xi-\overline{\xi})^2|\widehat{f}(\xi)|^2d\xi}{\int_{\mathbb{R}}|\widehat{f}(\xi)|^2d\xi}$$

be the variances of f in the space and frequency domains, $\overline{x}$ and $\overline{\xi}$ being the corresponding means. Then since f is bandlimited,

$$\int_{\mathbb{R}} \xi^2|\widehat{f}(\xi)|^2d\xi \leq (C\sqrt{d}\nu_j)^2 \int_{\mathbb{R}} |\widehat{f}(\xi)|^2d\xi\,,$$

and consequently, $\mathrm{Var}(\widehat{f}) \leq (C\sqrt{d}\nu_j)^2$ (the variance is always smaller than the second moment). The Heisenberg uncertainty principle implies that $\mathrm{Var}(f) \geq C'^2\nu_j^{-2}$.

As a conclusion, the extension operation satisfies a certain version of the Heisenberg principle relating the spectrum of the operator $\blacktriangle$ to the space and frequency localizations of the extensions of its eigenfunctions ϕ_j. This principle says that if $\blacktriangle\phi_j = \nu_j^2\phi_j$, then the operator $\mathbf{E}_B$ extends ϕ_j to a bandlimited function of band $\mathcal{O}(\nu_j)$ and localized in a tube of radius $\mathcal{O}(\frac{1}{\nu_j})$ around X. It is worthy to mention that similar results can be obtained for, say, Gaussian kernels.

4.3 Multiscale extension scheme

As a consequence of the previous section, where we have related the intrinsic and extrinsic Fourier analysis of a function on X, the size of the domain of extension and the bandwidth to which a given function f can be extended depends on int intrinsic spectrum. In particular, we know that each eigenfunction ϕ_j of Δ can be extended as a bandlimited function with bandwidth $C\nu_j^2$, at distance proportional to $\frac{1}{\nu_j}$ from X. This observation gives rise to a natural multiscale extension technique that we now describe.

Algorithm 2 (Multiscale extension scheme) *Fix a precision η and a condition number $\frac{1}{\delta}$.*

- *Precomputation phase: for each eigenfunction ϕ_j of Δ, compute the minimal frequency band B_j to which it can be extended using $\mathbf{E}_{B_j}$. This step can also be done by grouping the eigenfunctions in dyadic packets, and by computing a band for each packet.*
- *Extension phase: for any function f defined on X, compute its decomposition over $\{\phi_j\}$ and retain enough coefficients so that the relative error is of order η:*

$$f = \sum_{j \in S} \langle f, \phi_j \rangle_X \phi_j + \mathcal{O}(\eta \|f\|_X) \,,$$

and use the precomputed extensions of ϕ_j to extend f as

$$F = \sum_{j \in S} \langle f, \phi_j \rangle_X \mathbf{E}_{B_j} \phi_j \,.$$

This algorithm extends f as a linear combination of "atoms" with different localizations in time and frequency. More precisely, F is a sum of functions that oscillate at intrinsic frequency ν_j^2 on X and that vanish at distance $\frac{1}{\nu_j^2}$ from this set. From the implementation point of view, it is to be noted that in (4), the authors provide an efficient way to compute the eigenfunctions of Δ on a submanifold X.

We illustrate the space-frequency localization principle on figure 6 where different Fourier modes of the unit circle are extended to the plane using a Gaussian kernel.

5 conclusion

We have described the construction and properties of the geometric harmonics, and we have shown how they can be used to perform out-of-sample extension

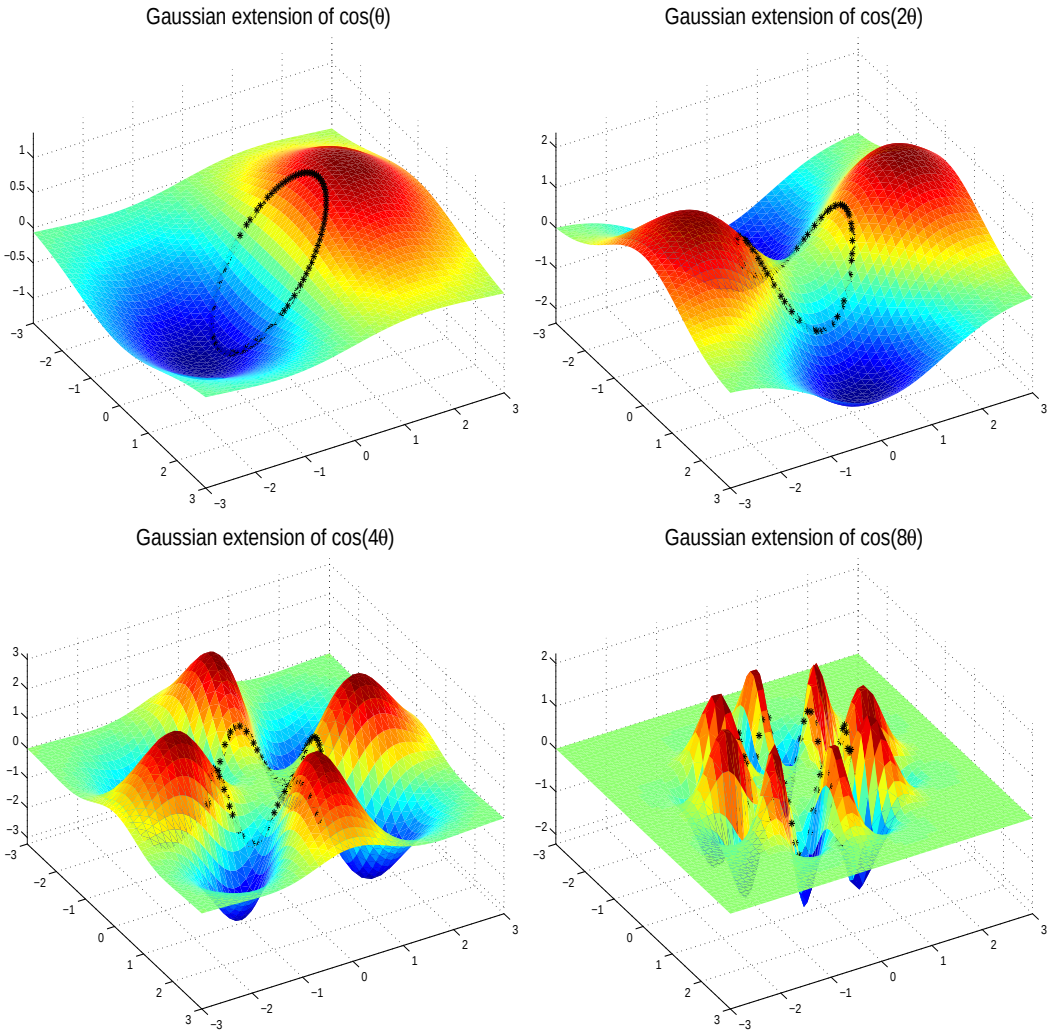

Fig. 6. Gaussian extensions of the functions $\cos(j\theta)$ for $j = 1, 2, 4, 8$, from the unit circle to the plane. Unlike the bandlimited extensions, these ones are much more localized in the plane, and they don't oscillate off the circle. These plots also illustrate the Heisenberg principle as functions with high frequencies are extended as functions with a small support.

of empirical function. The study of the case of submanifolds shows that these functions are an interesting tool for relating the intrinsic and extrinsic Fourier analyses.

6 Acknowledgements

We would like to thank Ann Lee, Mauro Maggioni and Naoki Saito for useful discussions and comments regarding this work.

A Reproducing kernel Hilbert spaces

In this section, we provide the basic background concerning reproducing kernel Hilbert spaces, and their connection to positive semi-definite kernels. A classical reference concerning this topic is (2).

Definition 15 (Reproducing kernel Hilbert space) *A space Hilbert $\mathcal{H}$ of functions defined on a set $\overline{X}$ is said to be a reproducing kernel Hilbert space if there exists a function ("kernel") $k : \overline{X} \times \overline{X} \to \mathbb{R}^n$ such that*

- *for almost every $x \in \overline{X}$, $k(x, \cdot) \in \mathcal{H}$,*
- *if $\langle \cdot, \cdot, \rangle_{\overline{X}}$ is the inner product in $\mathcal{H}$, then for all function $f \in \mathcal{H}$ and almost all $x \in \overline{X}$,*

$$\langle f, k(x, \cdot) \rangle_{\overline{X}} = f(x) \tag{A.1}$$

In (2), it is shown that the concepts of reproducing kernels and positive kernels are identical in the following sense: any reproducing kernel is positive semi-definite, and to any positive kernel k there corresponds a reproducing kernel Hilbert space $\mathcal{H}$ for which k is the reproducing kernel. The construction is given in (2).

In the particular case when $\overline{X} = \mathbb{R}^n$ and $k(x, y) = h(x - y)$, then more can be said. By Bochner's theorem, we know that the Fourier transform $\hat{h}$ of h is a finite positive measure. Assuming that this measure has the form $\hat{h}(\xi)d\xi$, then it can be checked that $\mathcal{H}$ is the space of functions f defined on $\mathbb{R}^n$ such that

$$\int_{\hat{h}(\xi)>0} |\hat{f}(\xi)|^2 \frac{d\xi}{\hat{h}(\xi)} < +\infty \,,$$

and such that $\hat{f}(\xi) = 0$ if $\hat{h}(\xi) = 0$. Consequently, this space is the Fourier transform of a weighted L^2 space, where the weight penalizes high frequencies since $\hat{h}$ is integrable.

In the case of the Bessel kernel, the function $\hat{h}$ is the indicator of a ball, and therefore $\mathcal{H}$ is imply a space bandlimited of bandlimited functions endowed with the classical inner product of $L^2(\mathbb{R}^n, dx)$.

B Expression of the Bessel kernel

In what follows, we derive the form of the kernel corresponding to functions whose Fourier transform is the indicator function of the ball of radius $\frac{B}{2}$ centered at the origin, namely:

$$k_B(x,y) = \int\limits_{\|\xi\|<\frac{B}{2}} e^{2i\pi\langle\xi,x-y\rangle} d\xi = \left(\frac{B}{2}\right)^{\frac{n}{2}} \frac{J_{\frac{n}{2}}(\pi B\|x-y\|)}{\|x-y\|^{\frac{n}{2}}}\,,$$

where J_ν is the Bessel function of the first kind of order ν. This kernel will be termed "Bessel kernel".

Since the kernel is really a function of $\|x-y\|$, we are looking for the form of the Fourier transform of the indicator of the unit ball in dimension n. To do so, we make use of a result known under the name of the Bochner-Coifman-Howe periodicity relations:

Lemma 16 *Let f be a radial function, and let $\mathcal{F}_n f(\xi) = h_n(\|\xi\|^2)$ be its Fourier transform in dimension n. Then the Fourier transforms of f in dimension n and $n+2$ are related in the following manner:*

$$h_{n+2}(u) = -\frac{1}{\pi} h_n'(u)\,.$$

In other words, to compute the Fourier transform $h_{n+2}(\|\xi\|^2)$ of f in $\mathbb{R}^{n+2}$, one can start from the Fourier transform $h_n(\|\xi\|^2)$ in dimension n, view this function as a function of $\|\xi\|^2$ and compute its derivative in this variable.

PROOF. Since any radial function or tempered distribution can be approximated as a sum of Gaussians, one merely needs to verify the relation for $f(x) = e^{-\alpha r^2}$. In this case,

$$\mathcal{F}_n f(\xi) = \left(\frac{\pi}{\alpha}\right)^{\frac{n}{2}} e^{-\frac{\pi^2\xi^2}{\alpha}}\,.$$

Thus

$$h_n(u) = \left(\frac{\pi}{\alpha}\right)^{\frac{n}{2}} e^{-\frac{\pi^2 u}{\alpha}}\,,$$

and the identity is satisfied. □

Using this lemma we can now conclude:

Proposition 17 *In dimension n, the Bessel kernel has the following form:*

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

Moreover, if n is odd, then the simpler formula can be used:

$$k_B(x,y) = M_B \left(\frac{1}{r}\frac{d}{dr}\right)^{\frac{n-1}{2}} sinc(Br)\,,$$

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

$$M_c = \left(\frac{B}{2}\right)^{\frac{n}{2}} \sqrt{2B}(-1)^{\frac{n-1}{2}}\,.$$

PROOF. By a trivial scaling argument, we may assume that $B = 2$.

Then if $n = 1$, then

$$k(x,y) = \int_{-1}^{1} e^{2i\xi\pi(x-y)}d\xi = 2\mathrm{sinc}(2\|x-y\|) = \frac{J_{\frac{1}{2}}(2\pi\|x-y\|)}{\|x-y\|^{\frac{1}{2}}}\,,$$

where the third equality is obtained using 10.1.1 and 10.1.11 in (1).

If $n = 2$, then in polar coordinates (ρ,θ):

$$\begin{aligned}
\int_{\|\xi\|<1} e^{2i\pi\langle\xi,x-y\rangle}d\xi &= \int_0^1\int_0^{2\pi} e^{2i\pi r\rho\cos\theta}d\theta\rho d\rho\,,\\
&= 2\pi\int_0^1 J_0(2\pi r\rho)\rho d\rho \text{ by 9.1.21 in (1)},\\
&= \frac{1}{2\pi r^2}\int_0^{2\pi r} uJ_0(u)du\,,\\
&= -\frac{1}{r}J_0'(2\pi r) \text{ since by 9.1 in (1) } (uJ_0(u))' = -uJ_0(u)\,,\\
&= \frac{J_1(2\pi r)}{r} \text{ by 9.1.28 in (1)}.
\end{aligned}$$

For higher orders we proceed by induction on n, noting that if

$$h(u) = \frac{J_{\frac{n-2}{2}}(2\pi\sqrt{u})}{u^{\frac{n-2}{4}}},$$

then

$$\begin{aligned}
h'(u) &= \frac{J'_{\frac{n-2}{2}}(2\pi\sqrt{(u)})\frac{\pi}{\sqrt{(u)}}u^{\frac{n-2}{4}} - J_{\frac{n-2}{2}}(2\pi\sqrt{u})\frac{n-2}{4}u^{\frac{n-2}{4}-1}}{u^{\frac{n-2}{2}}}, \\
&= \frac{2\pi\sqrt{u}J'_{\frac{n-2}{2}}(2\pi\sqrt{u}) - \frac{n-2}{2}J_{\frac{n-2}{2}}(2\pi\sqrt{u})}{2u^{\frac{n-2}{4}+1}}, \\
&= -\frac{\pi\sqrt{u}J_{\frac{n}{2}}(2\pi\sqrt{u})}{u^{\frac{n-2}{4}+1}} \text{ according 9.1.27 in (1)}, \\
&= -\frac{\pi J_{\frac{n}{2}}(2\pi\sqrt{u})}{u^{\frac{n}{2}}}.
\end{aligned}$$

Now invoking lemma 16 yields the result. Finally, to obtain a formula in terms of the variable r instead of r^2, notice that $d(r^2) = rdr$, and this implies that for odd values of n

$$k(x,y) = 2(-1)^{\frac{n-1}{2}}\left(\frac{1}{r}\frac{d}{dr}\right)^{\frac{n-1}{2}}\text{sinc}(2r).$$

□

References

[1] M. Abramowitz and I. Stegun, "Handbook of mathematical functions", Dover, 1965.

[2] N. Aronszajn, "Theory of reproducing kernels", *Transactions of the American Mathematical society*, Vol. 68, No. 3, May 1950.

[3] Y. Bengio, J-F. Paiement and P. Vincent, "Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering", *Technical report 1238*, Universitde Montréal, July2003.

[4] R.R. Coifman and S. Lafon, "Diffusion maps", To appear.

[5] G.B. Folland, "Introduction to partial differential equations", *Princeton University Press*, 1976.

[6] C. Fowlkes, S. Belongie, and J. Malik, "Efficient Spatiotemporal Grouping Using the Nystrm Method", *IEEE Computer Vision and Pattern Recognition*, December 2001.

[7] W.H. Press, S.A. Teulkosky, W.T. Vetterling and B.P. Flannery, "numercial recipes in C", *Cambridge University Press*, chapter 18.1.

[8] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty I", *The Bell System technical journal*, Vol. 40, pp 43-64, January 1961.

[9] 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*, Vol. 43, pp 3009-3058, November 1964.

[10] C. Williams and M. Seeger, "Using the Nyström method to speed up kernel machines", NIPS 2000.