A Laplacian based graph out of sample extension

Nathanael Perraudin and Pierre Vandergheynst

January 22, 2017

Abstract

TBD

Contents

1 Introduction

A new perspective Let us forget about the riemanian manifold for a moment and consider that we have a probability distribution in $\mathbb{R}^d$ with measure $d\mu(y)$, i.e:

$$\int_{\mathbb{R}} d\mu(y) = 1.$$

For a kernel $k(x, y)$, we define the degree as

$$d(y) = \int_{\mathbb{R}} k(x, y) d\mu(x)$$

Let us define the normalized kernel $h(x, y) = \frac{k(x,y)}{\sqrt{d(x)d(y)}}$. We are interested in the following operator

$$\mathcal{L}' f(y) = f(y) - \int_{\mathbb{R}} h(x, y) f(x) d\mu(x).$$

The lowest eigenvector of $\mathcal{L}'$ form a normalized embedding in lower dimension of the data???

Let us sample n data points. We obtain the Matrices $\boldsymbol{K}$ and $\boldsymbol{H}$. Let us suppose that $\boldsymbol{f}$ is the vector of signals values sample points. Let us denote $\boldsymbol{h}_y[i] = h(x_i, y)$. Our estimator of f becomes

$$\dot{f}(y) = \langle \boldsymbol{h}_y, \boldsymbol{H}^{-1} \boldsymbol{f} \rangle$$

Plan

- We investigate ways to extend a signal s out of the graph from a finite number n of samples $s(x_i) = \boldsymbol{s}[i]$. The graph is generated from a kernel $k(x_i, x_j)$ that asses how close x_i and x_j are. The point x_i are typically sampled from a Manifold $\mathcal{M}$. The solution has be

  1. **Stable:** We would like that its behavior is well controlled in practice. *Additionally since we do not know the Manifold $\mathcal{M}$, we would like that the extension has a value 0 where x is in probability not in the manifolds.*
  2. **Efficient:** We would like to be able to compute it with a minimum number of computation even when the number of samples is large.
  3. **Theoretically justified:** We would like to understand what is happening.
  4. **Consistent:** We would like that it follows the graph shape and preserve the graph properties.

- Here are few solutions:

  1. The first idea would be to add a single node to the graph and to solve an in-painting problem. The explicit solution this problem would be

  $$s(x) = \frac{\sum_{i=1}^n k(x_i, x) s(x_i)}{\sum_{i=1}^n k(x_i, x)}, \tag{1}$$

  which is nothing but a weighted average of the sample $s(x_i)$. This simple solution has two drawback. First, it does not take into account of the Manifold and thus is not 0 where in probability there is no manifold. Second, it ignores completely the graph structure and is thus not consistent.

  2. One idea would be to use the Nystrom approximation of the Laplacian to extend the Fourier basis. In this case, we would take into account of graph structure. This idea has two drawbacks. It is not stable as we cannot extend the eigenvector associated with eigenvalue 0. It is computationally expensive for large graph.

- Instead, our solution is based on spectral decomposition of the point-cloud Laplacian operator. It has been studied in [15]. There are two important theorems that characterize the eigenvectors of the point-cloud Laplacian on a manifold $\mathcal{M}$.
  - Let us define the degree $d(x) = \sum_{i=1}^{n} k(x_i, x)$ and define the band

    $$\mathrm{rg}(d) = [\min_{x \in \mathcal{M}} d(x), \max_{x \in \mathcal{M}} d(x)],$$

    then all eigenvectors with associated eigenvalue stratifying $\lambda_\ell \notin \mathrm{rg}(d)$ can be extended on the manifold $\mathcal{M}$. In the case of normalized Laplacian, the degree is 1.
  - The eigenvectors are defined that way because they contains the factor $\frac{1}{\lambda_\ell - d}$ that diverge in case $d \to \lambda_\ell$
  - In practice we do not know the manifold and if we ask for a point far from the graph, the degree will be almost 0. This implies that for the combinatorial case, we cannot extend eigenvector far away from the graph. In the normalized case, we can extend almost (unless $\lambda = 1$) all eigenvectors everywhere.
  - The normalized out of sample extension can thus be defined as follow:
    1. Compute the graph Fourier transform of $\boldsymbol{s}$
    2. Use the Fourier coefficients to weights the extended graph eigenvectors
  - From our derivation, we derive a computationally efficient way to extend the signal that does not involve an eigenvalue decomposition of the Laplacian.
  - The frequency associated with eigenvalue $\lambda = 1$ can not be reconstructed! As a result, 1 is a kind of cutoff frequency.
  - This approach does not work for the combinatorial case.
- As a second part of the work, based on a similar framework, we extend the graph localization operator to the manifold
  - Instead of being characterized by $\frac{1}{\lambda_\ell - d}$, the obtained atoms depends only on $\frac{g\lambda_\ell - g(d)}{\lambda_\ell - d}$.
  - Providing that the derivative of g is controlled, we can extend the atoms by continuity at places where the eigenvectors are not defined. The resulting atoms are well defined in the entire space for both the normalized and the combinatorial Laplacian.
  - We prove then that the extended atoms are continuous only if $g(d_i) = 0$ (respectively $g(1) = 0$), where d_i is the degree of node i. Note that this is way better that to ask $g(\mathrm{rg}(d)) = 0$.
  - This also suggest a frequency limit, that is however much better in the combinatorial case.
  - In order to extend a signal, we can thus express it as a sum of atoms that we will extend separately.
  - We finally prove that both approaches are equivalent in the normalized case providing that $g(d) = 1$. In the combinatorial case, I have to finish the theory.
- In general, the normalized case seems to behave way better.

1.1 The problem

In this note, we define a new object that allows the extension of a kernel defined on a graph over an underlying space or manifold. We try to build the theoretical basis for the extension and define some of its main properties.

The extension is based on the point-cloud Laplacian operator that is a generalization of the Graph Laplacian. In previous work, this object is only used to prove convergence between the graph and the Laplace-Beltrami Laplacian[2, 3]. Thanks to the extension, we leverage the use of the point-cloud Laplacian to extend a function (or a kernel) out of the graph nodes. This work has two main goals. First, we would like to handle new samples without inserting them inside the graph, making it possible to use the graph structure without expensive computation. Second we would like to establish a theoretical consistency results between the spectral atoms of graphs and Manifold.

The following points summaries the general idea of this contribution:

1. Define a "spectral" basis for the point cloud Laplacian using its eigen-decomposition.
2. Define the atoms by localizing kernel utilizing the aforementioned basis.
3. Write a function on the Manifold as a sum of these atoms

Goal: extend the spectral graph technique to manifold without reconstructing the graph! Theoretically, we need to answer the following questions:

1. What are the basic consistency requirement for the OOSE?
2. In the large sample limit, does the extended function converges toward an underlying function defined on the Manifold?
3. If it converges what are the properties of this function?

1.2 Prior works

Our contribution is mainly constructed on the shoulders of four important publications. The story begins with Belkin and Niyogi in [4]. This contribution explains how to benefit from graphs in manifold learning problem based on the idea that graphs estimate the manifold. Later, in [2] they go further and prove the convergence of the graph Laplacian operator converges toward the Laplace Beltrami operator. This result is based on the heat diffusion and motivates the use of the Gaussian function to compute the weights of the graph from the distances. In a more theoretical contribution, they also prove the convergence of the eigenvectors and the eigenvalues [3]. This final step is possible thanks to spectral clustering convergence result present in [15].

- spectral graph frames [9] [13] [14] [11]
- wavelet for manifolds [8]

TBD: other less important publications

1.3 Definitions and nomenclature

We now introduce the main setting of the paper as well as the basic definitions and hypotheses. We work with 2 different spaces: 1) A continuous space defined by a manifold $\mathcal{M}$ embedded in $\mathbb{R}^N$ and 2) the discrete space of the graph vertices corresponding to n points sampled from $\mathcal{M}$. Variables related to the discrete space will be marked in bold font, e.g $\boldsymbol{f} \in \mathbb{R}^n$. Throughout this paper, the calligraphic typology is generally reserved for object related to the manifold $f \in \mathcal{C}(\mathcal{M})$. We use the standard notation f for signal object related to the point-cloud laplacian operator (defined bellow). To further help the reader, we also differentiate a continuous function $f : \mathcal{M} \to \mathbb{R}$ from a discrete signal $\boldsymbol{f} : \mathcal{V} \to \mathbb{R}$ with the use of parenthesis $(\cdot)$ and bracket $[\cdot]$. The function $\rho_n : \mathcal{C}(\mathcal{M}) \to \mathbb{R}^n$ restricts the continuous function f on the sampled point, i.e: $f(x_i) = \big(\rho_n(f)\big)_i = \boldsymbol{f}[i]$. We use the subscript n and the superscript t when we want to emphasis that a quantity depends on the number of samples n and the kernel parameter t. Finally the prime sign is used for the normalized case.

The construction presented in this contribution can also be done using to the following operator $\boldsymbol{L}'' = \boldsymbol{D}^{-1}\boldsymbol{L}$. For simplicity, we do not present it.

Manifold Laplacian We consider a compact smooth Manifold $\mathcal{M}$ embedded in a Euclidean space $\mathbb{R}^N$. This embedding induce a measure μ on the manifold. For example, the measure of a closed curve, a 1-dimensional Manifold, is the usual curve length.

Given $\in \mathcal{M}$, the tangent space $T_p\mathcal{M}$ can be identified by the affine space of tangent vectors to $\mathcal{M}$ at point . For a differentiable function $f : \mathcal{M} \to \mathbb{R}$, we denote $\nabla_{\mathcal{M}} f$ the gradient field on the manifold. Please recall that $\nabla_{\mathcal{M}} f(p)$ is a vector in the tangent space $T_p\mathcal{M}$ pointing into the direction of the fastest ascent of f at point p.

The Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ can be defined as the divergence of the gradient[1]:

$$\Delta_{\mathcal{M}} f = -\text{div}_{\mathcal{M}}(\nabla_{\mathcal{M}} f)$$

Alternatively, the Laplace Beltrami operator could be defined as the only operator satisfying for two differentiable functions f and g:

$$\int_{\mathcal{M}} f(x)\Delta_{\mathcal{M}} g(x) d\mu(x) = \int_{\mathcal{M}} \langle \nabla_{\mathcal{M}} f, \nabla_{\mathcal{M}} g \rangle d\mu(x)$$

In a Euclidean space, the Laplace Beltrami operator is the ordinary Laplacian, i.e:

$$\Delta f = \sum_{i=1}^{k} \frac{\partial^2 f}{\partial x_i^2}.$$

I'm not sure the notation is consistent here. On a k-dimensional manifold $\mathcal{M}$ ($k \leq N$), in a local coordinate system $(x_1, \ldots, x_k)$, with a metric tensor g_{ij}, the Laplace-Beltrami operator applied to a function $f(x_1, \ldots, x_k)$ is given by:

$$\Delta_{\mathcal{M}} f = \frac{1}{\sqrt{\det(g)}} \sum_j \frac{\partial}{\partial x_j} \left(\sqrt{\det(g)} \sum_i g^{ij} \frac{\partial f}{\partial x_i} \right)$$

where g^{ij} are the components of the inverse of the metric tensor G: G^{-1}.

$\Delta_{\mathcal{M}}$, as a positive semi definite operator on $\mathcal{C}^\infty(\mathcal{M})$, possesses an orthonormal basis of eigenfunctions denoted $\{u_\ell 0 \leq \ell \infty\}$ where $\forall \lambda_\ell, u_\ell \in \mathcal{C}^\infty(\mathcal{M})$. For consistency, we order the u_ℓ such that the corresponding eigenvalues form a non-decreasing sequence. In this case, u_0 is constant, $\lambda_0 = 0$ and all other $\lambda_\ell > 0$.

This orthonormal basis is complete with respect of the space of function $\mathcal{C}^\infty(\mathcal{M})$, i.e: a smooth function $f \in \mathcal{C}^\infty(\mathcal{M})$ can be uniquely decomposed as $f = \sum_{\ell=0}^{\infty} a_\ell u_\ell$. Additionally, $\Delta_{\mathcal{M}} f = \sum_{\ell=0}^{\infty} \lambda_\ell a_\ell u_\ell$. By extension, we call the basis u_ℓ the Manifold Fourier basis and the a_ℓ the Manifold Fourier coefficients.

Graph Laplacian: Given n samples of $\mathcal{M}$: $x_1, \ldots x_n$, we construct a complete weighted graph. Each node correspond to a sample. Each pair of node is connected with a weight $\boldsymbol{W}_{i,j} = w(x_i, x_j) = \frac{1}{n} k(x_i, x_j)$, where $k : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_+^*$ is a kernel. We restrict ourselves to strictly positive kernel such as $k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{4t}}$. The sum of the weights gives the degree $\boldsymbol{d}[i] = \sum_{j=1}^{n} w(x_i, x_j)$ and the diagonal degree matrix $\boldsymbol{D}$ such that $\boldsymbol{D}_{ii} = \boldsymbol{d}[i]$. Using our definitions, the degrees can be arbitrarily close to $\frac{1}{n}$ but they are always strictly greater than $\frac{1}{n}$. Note that the addition of self-loop, i.e: $w_{x_i, x_i} = \frac{1}{n}$. It will play an important role in the extension.

The graph Laplacian operator is defined as

$$\boldsymbol{L}\boldsymbol{f}[i] := \frac{1}{n} \left(\boldsymbol{f}[i] \sum_{j=1}^{n} k(x_i, x_j) - \sum_{j=1}^{n} \boldsymbol{f}[j] k(x_i, x_j) \right) \tag{2}$$

Note that this operator acts only on the graph points. The associated graph Laplacian matrix becomes $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$. In practice, other definition of the Laplacian are used. In this contribution, we will extend our results for the normalized Laplacian

$$\boldsymbol{L}'\boldsymbol{f}[i]) = \boldsymbol{f}[i] - \frac{1}{\sqrt{\boldsymbol{d}[i]}} \sum_{j=1}^{n} \boldsymbol{f}[i] \frac{k(x_i, x_j)}{n\sqrt{\boldsymbol{d}[j]}}, \tag{3}$$

[1] Modulo a minus sign coming from the definition of the divergence

with the more convenient matrix form:

$$\boldsymbol{L}' = \boldsymbol{I} - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{D}^{-\frac{1}{2}} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}}.$$

Those operators are central in graph signal processing [13, 6].

The graph Laplacian is a positive semi definite operator that possess a full orthonormal eigen-decomposition written as:

$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$$

The matrix $\boldsymbol{U}$ is called the graph Fourier matrix. Similarly to the manifold case, we order the eigenvectors $\boldsymbol{u}_\ell$ such that their corresponding eigenvalues $\boldsymbol{\lambda}_\ell$ are in a non-decreasing order.

Point cloud Laplacian operator To prove that the graph Laplacian operator converge to the Laplace-Beltrami operator, Belkin and Niogi define the point cloud Laplacian operator in [2]. It is a generalization of the graph Laplacian operator that apply on a function $f : \mathcal{M} \to \mathbb{R}$:

$$Lf(y) := \frac{1}{n}\left(f(y)\sum_{j=1}^{n} k(y, x_j) - \sum_{j=1}^{n} f(x_j)k(y, x_j) \right) \tag{4}$$

We observe that $Lf(y)$ depends only of the values of f at $y, x_1, x_2, \ldots, x_n$. It is thus independent from the other points of the manifold, an essential property that will be kept by our extension. Let us select $y \in \mathcal{M}$, we use the following matrix notation:

$$L = \begin{pmatrix} \boldsymbol{L} & 0 \\ -\boldsymbol{w}_y & d_y \end{pmatrix}$$

with $\boldsymbol{w}_y[j] = w(y, x_j)$ and $d_y = \sum_{j=1}^{N} w(y, x_j)$. The matrix L is not symmetric. It reflects the fact that the signal on the graph has an influence on $Lf(y)$ but $f(y)$ has no influence on what on the graph $Lf(x_i)$.

In [15], we also find a normalized version of the point-cloud Laplacian operator. For a function $f : \mathcal{M} \to \mathbb{R}$:

$$L'f(y) := f(y) - \frac{1}{\sqrt{d_y}}\sum_{j=1}^{n} f(x_j)\frac{k(y, x_j)}{n\sqrt{\boldsymbol{d}[j]}} \tag{5}$$

with the matrix form

$$L' = D^{-\frac{1}{2}}LD^{-\frac{1}{2}} = \begin{pmatrix} \boldsymbol{L}' & 0 \\ -d_y^{-\frac{1}{2}}\boldsymbol{w}_y\boldsymbol{D}^{-\frac{1}{2}} & 1 \end{pmatrix}.$$

Graphs atoms through localization operator Using the spectral decomposition of the Laplacian, one can define atoms on the graphs. This has been done in [9, 14]. The atom generated by the kernel $g : \mathbb{R}_+ \to \mathbb{R}$ and localized at node i is defined as:

$$\boldsymbol{T}_i g[j] := \sum_{\ell=0}^{N-1} g(\boldsymbol{\lambda}_\ell)\overline{\boldsymbol{u}_\ell}[i]\boldsymbol{u}_\ell[j] = \left(g(\boldsymbol{L})\right)_{ij}, \tag{6}$$

where $\overline{\boldsymbol{u}_\ell}$ is the complex conjugate of $\boldsymbol{u}_\ell$. For a sufficiently regular function g, it has been proved in [14] that this operation localizes the filter g around the vertex i.

Manifold atoms Similar atoms have been defined and studied on manifolds in [8] as:

$$\mathcal{T}_{\mathcal{M}x}g(y) = \sum_{\ell=0}^{\infty} g(\lambda_\ell)\boldsymbol{u}_\ell(x)\bar{\boldsymbol{u}}_\ell(y)$$

Both atoms have a very similar definition. The resemblance arises from the fact they are both defined spectrally.

From the observation, we ask ourself the following question. Since it has been proved that $\boldsymbol{u}_\ell[i] \to u_\ell(x_i)$ as $n \to \infty$, does $\boldsymbol{T}_i g[j] \to \mathcal{T}_{\mathcal{M}x_i}g(x_j)$? The answer is yes and is one result of the paper.

1.4 Existing results

Link between graph and Manifold Theorem 3.1 of [2] shows the link between $\boldsymbol{L}$ and $\Delta_{\mathcal{M}}$

Theorem 1 (Laplacian convergence, Belkin/Nyogi)**.** *Let data points $x_1, \ldots, x_n$ be sampled from a uniform distribution on a manifold $\mathcal{M} \subset \mathbb{R}^N$. Using the kernel $k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{4t}}$, define a sequence $t(n) = n^{-\frac{1}{k+2+\alpha}}$, where $\alpha > 0$ and let $f \in \mathcal{C}^\infty(\mathcal{M})$, then the following holds:*

$$\lim_{n\to\infty} \frac{1}{t(4\pi t_n)^{\frac{k}{2}}} L_n^{t_n} f(x) = \frac{1}{\mathrm{vol}(\mathcal{M})} \Delta_{\mathcal{M}} f(x)$$

where the limit is taken in probability and $\mathrm{vol}(\mathcal{M})$ is the volume of the manifold with respect to the canonical measure.

Additionally Theorem 2.1 of [3] shows the convergence between the eigenfunction and the eigenvalues of both Laplacians.

Theorem 2 (Laplacian spectral convergence, Belkin/Nyogi)**.** *Using $k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{4t}}$, let $\boldsymbol{\lambda}_{\ell,n}^t$ be the i^{th} eigenvalue of $\frac{1}{t(4\pi t)^{\frac{k}{2}}} L_n^t$ and $u_{\ell,n}^t$ be the corresponding eigenfunction (which, for each fixed ℓ exist for t sufficiently small). Let λ_ℓ and u_ℓ be the corresponding eigenfunction of $\frac{1}{\mathrm{vol}(\mathcal{M})}\Delta_{\mathcal{M}}$ respectively. Then there exists a sequence $t(n) \to 0$, such that*

$$\lim_{n\to\infty} \boldsymbol{\lambda}_{\ell,n}^t = \lambda_\ell$$

and

$$\lim_{n\to\infty} \|u_{\ell,n}^t(x) - u_\ell(x)\|_2 = 0$$

where the limits are in probability.

We also need Proposition 22 of [15]

Theorem 3 (Spectrum of L_n)**.**

1. *If $u \in C(\mathcal{M})$ is a (continuous) eigenfunction of L_n with eigenvalue λ, then the vector $\rho_n(u_n) = \boldsymbol{u}_n \in \mathbb{R}^n$ is an eigenvector of the matrix $\boldsymbol{L}_n$ with eigenvalue λ.*
2. *Let $\lambda_\ell \notin \mathrm{rg}(\boldsymbol{d})$, be an eigenvalue of L_n with eigenfunction $u_n \in C(\mathcal{M})$, then the vector u_n is of the form*

$$u_\ell(y) = \frac{\sum_{i=1}^n k(y, x_i)\boldsymbol{u}_\ell[i]}{d_y - \lambda_\ell} \tag{7}$$

3. *If $\boldsymbol{u}_\ell$ is an eigenvector of the matrix $\boldsymbol{L}_n$ with eigenvalue $\lambda \in \mathrm{rg}(\boldsymbol{d})$, then u_ℓ defined in* (7) *is an eigenfunction of L_n with eigenvalue λ.*
4. *The essential spectrum of L_n coincides with the range of the degree function, that is, $\sigma_{\mathrm{ess}}(L) = \mathrm{rg}(d)$. All eigenvalues of L_n are nonnegative and can have accumulation points only in $\mathrm{rg}(d)$. The analogous statement hold when $n \to \infty$.*

Proposition 9 of [15] corresponds to an analogous theorem for the corresponding normalized operator: L_n'

Theorem 4 (Spectrum of L_n')**.**

1. *If $u \in C(\mathcal{M})$ is a (continuous) eigenfunction of L_n' with eigenvalue λ, then the vector $\rho_n(u_n) = \boldsymbol{u}_n \in \mathbb{R}^n$ is an eigenvector of the matrix $\boldsymbol{L}_n'$ with eigenvalue λ.*
2. *Let $\lambda \neq 1$, be an eigenvalue of L_n' with eigenfunction $u_n \in C(\mathcal{M})$, then the vector u_n is of the form*

$$u_\ell(y) = \frac{\sum_{i=1}^n k(y, x_i)\boldsymbol{u}_\ell[i]}{1 - \lambda} \tag{8}$$

   check what is k? is it normalized

3. *If $\boldsymbol{u}_\ell$ is an eigenvector of the matrix $\boldsymbol{L}_n'$ with eigenvalue $\lambda \neq 1$, then u_ℓ defined in* (8) *is an eigenfunction of L_n' with eigenvalue λ.*
4. *The essential spectrum of L_n' consists of finitely many nonnegative eigenvalues with finite multiplicity. The essential spectrum of L_n'*

define the essential spectrum!!!

2 Theoretical Exposition

2.1 Out of sample extension definition

For a continuous kernel $g : \mathbb{R}_+ \to \mathbb{R}$, the localized atoms on the graph are the columns of the following matrix.

$$g(\boldsymbol{L}) = \boldsymbol{U} g(\boldsymbol{\Lambda}) \boldsymbol{U}^*$$

In order to define extend atoms from the graph to the manifold, we use a similar definition based on the point cloud Laplacian L instead of $\boldsymbol{L}$. In order to do so, we first need to explore the spectral properties of the point cloud graph Laplacian L. As a result, we propose a variation of Theorems 3 and 4.

Theorem 5. *For $d_y \neq \boldsymbol{\lambda}_\ell$ and $\ell = 0, 1, \ldots, N-1$, the operator $L = \begin{pmatrix} \boldsymbol{L} & 0 \\ -\boldsymbol{w}_y & d_y \end{pmatrix}$ can be decomposed as*

$$L = U \Lambda U^{-1} = \begin{pmatrix} \boldsymbol{U} & 0 \\ \boldsymbol{c} & 1 \end{pmatrix} \begin{pmatrix} \Lambda & 0 \\ \boldsymbol{0} & d_y \end{pmatrix} \begin{pmatrix} U^* & 0 \\ -cU^* & 1 \end{pmatrix}$$

with $\boldsymbol{c} = \boldsymbol{w}_y \boldsymbol{U} (d\boldsymbol{I} - \boldsymbol{\Lambda})^{-1}$

Theorem 6. *For $d_y \neq 1$, the operator $L' = \begin{pmatrix} \boldsymbol{L}' & 0 \\ -d_y^{-\frac{1}{2}} \boldsymbol{w}_y \boldsymbol{D}^{-\frac{1}{2}} & 1 \end{pmatrix}$ can be decomposed as*

$$L' = U' \Lambda' U'^{-1} = \begin{pmatrix} \boldsymbol{U}' & 0 \\ \boldsymbol{c} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda}' & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{U}'^* & 0 \\ -\boldsymbol{c}\boldsymbol{U}'^* & 1 \end{pmatrix}$$

with $\boldsymbol{c} = d_y^{-\frac{1}{2}} \boldsymbol{w}_y \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{U}' (\boldsymbol{I} - \boldsymbol{\Lambda}')^{-1}$

The proofs are given in Appendix B.

Example 1 (Eigenvectors of the grid)**.** *Let us analyze a regularly spaced point cloud of* 10×10 *points. In Figure 1... TBD*

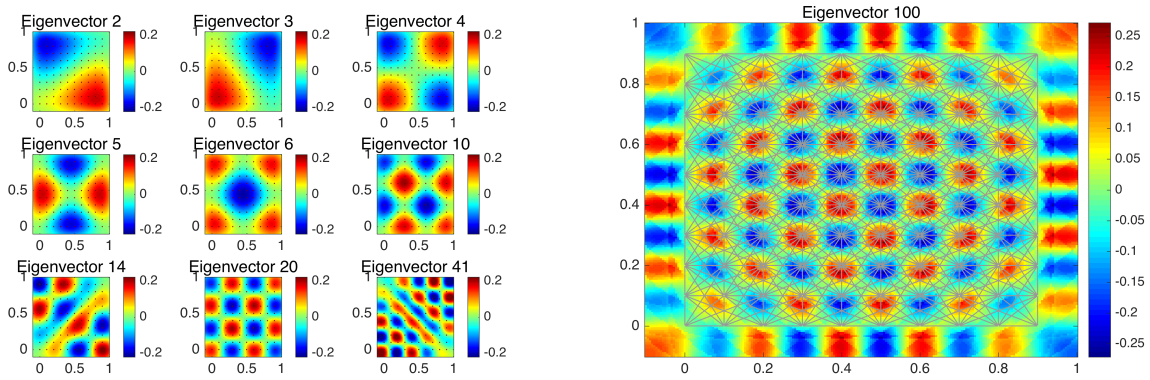

Figure 1: Extended eigenvectors for the grid graph. The graph is plotted on the right. Due to the symmetric shape of the data. The eigenvectors are similar to classical Fourier modes.

Example 2 (Eigenvectors of the spiral data)**.** *In Figure 2... TBD*

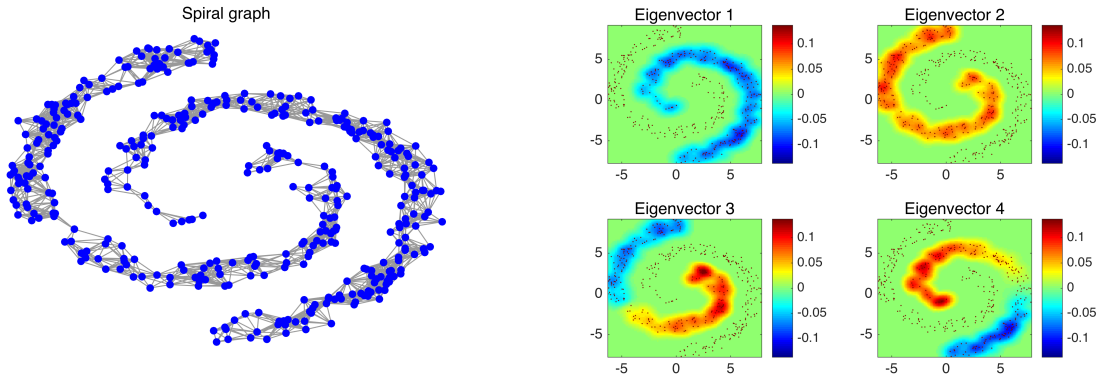

Figure 2: Extended eigenvectors for the spiral graph. Left: graph. Right: 4 smallest eigenvectors. They follow the shape of the graph. The two first eigenvectors are indicative functions for the clusters.

Based on this decomposition, we define a point cloud Laplacian spectral operator $g(L)$ and $g(L')$ as:

$$\begin{aligned} g(L) &= Ug(\Lambda)U^{-1} \\ &= \begin{pmatrix} \boldsymbol{U} & 0 \\ \boldsymbol{c} & 1 \end{pmatrix} \begin{pmatrix} g(\boldsymbol{\Lambda}) & 0 \\ 0 & g(d) \end{pmatrix} \begin{pmatrix} \boldsymbol{U}^* & 0 \\ -\boldsymbol{c}\boldsymbol{U}^* & 1 \end{pmatrix} \\ &= \begin{pmatrix} g(\boldsymbol{L}) & 0 \\ \boldsymbol{w}_y\boldsymbol{U}(d_y\boldsymbol{I}-\boldsymbol{\Lambda})^{-1}(g(\boldsymbol{\Lambda})-g(d_y)\boldsymbol{I})\boldsymbol{U}^* & g(d_y) \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} g(L') &= U'g(\Lambda)U'^{-1} \\ &= \begin{pmatrix} g(\boldsymbol{L}') & 0 \\ d_y^{-\frac{1}{2}}\boldsymbol{w}_y\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}'(\boldsymbol{I}-\boldsymbol{\Lambda}')^{-1}\left(g(\boldsymbol{\Lambda}')-g(\boldsymbol{I})\right)\boldsymbol{U}'^* & g(1) \end{pmatrix} \end{aligned}$$

It is important to note that for some y, U or U' are not defined but $g(L)$ and $g(L')$ exist if $\frac{g(d_y)-g(\lambda)}{d-\lambda}$, $\frac{g(1)-g(\lambda)}{d-\lambda}$ respectively are defined for every λ being an eigenvalue of the graph. To ensure that this quantity exists, we could require that the derivative of g is finite for all graph eigenvalues.

We are now ready to define an out of sample localization operator for a kernel defined in the graph spectral domain (a graph filter). It generalizes the graph localization operator $\boldsymbol{T}_i g$.

Definition 1 (Graph out of sample localization operator)**.** *For a derivable function g, the graph out of sample localization operator at point y is defined as:*

$$T_i g(y) := \left[\boldsymbol{w}_y\boldsymbol{U}(d_y\boldsymbol{I}-\boldsymbol{\Lambda})^{-1}(g(\boldsymbol{\Lambda})-g(d_y)\boldsymbol{I})\boldsymbol{U}^*\right]_i = \left[\boldsymbol{w}_y h_y(\boldsymbol{L})\right]_i = \left[h_y(\boldsymbol{L})\boldsymbol{w}_y^*\right]_i$$

where $h_y(x) = \lim_{z\to x} -\frac{g(d_y)-g(z)}{d_y-z}$.

Alternatively, we can write the Localization operator as

$$T_i g(y) := \sum_{\ell=1}^{n} g(\lambda_\ell)u_\ell(y)\boldsymbol{v}_\ell[i] \tag{9}$$

where $u_\ell(y) = [\boldsymbol{w}_y \boldsymbol{U}(d_y \boldsymbol{I} - \boldsymbol{\Lambda})^{-1}]_\ell$ and

$$\boldsymbol{v} = \boldsymbol{w}_y \boldsymbol{U}(\boldsymbol{\Lambda} - d_y \boldsymbol{I})^{-1} \boldsymbol{U}^* = \boldsymbol{w}_y (\boldsymbol{L} - d_y \boldsymbol{I})^{-1}.$$

Definition 2 (Normalized graph out of sample localization operator)**.** *For a derivable function g, the normalized graph out of sample localization operator at point y is defined as:*

$$T'_i g(y) := \left[d_y^{-\frac{1}{2}} \boldsymbol{w}_y \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{U}' (\boldsymbol{I} - \boldsymbol{\Lambda}')^{-1} \left(g(\boldsymbol{\Lambda}') - g(\boldsymbol{I}) \right) \boldsymbol{U}'^* \right]_i = \left[d_y^{-\frac{1}{2}} \boldsymbol{w}_y \boldsymbol{D}^{-\frac{1}{2}} h_y(\boldsymbol{L}) \right]_i = \left[h_y(\boldsymbol{L}) \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{w}_y^* d_y^{-\frac{1}{2}} \right]_i$$

where $h_y(x) = \lim_{z \to x} -\frac{g(1)-g(z)}{1-z}$.

Example 3 (Atoms on the spiral data)**.** *In Figure 3, TBD*

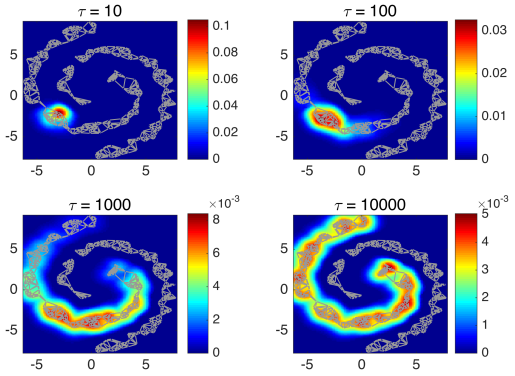

Figure 3: Extended atoms for the spiral graph. The kernel is $e^{-\frac{\tau}{\lambda_{\max}} x}$.

Definition 3 (Analysis operator)**.** *Given a graph filter-bank $\{g_k\}_{k \in 1 \dots K}$ and a function $f : \mathcal{M} \to \mathbb{R}$, the "out of sample" analysis $\mathbf{A}_{\mathcal{M},g} : \mathcal{M} \to \mathbb{R}^{nK}$ is defined as*

$$c[i,k] = \mathbf{A}_{\mathcal{M},g} f[i,k] := \langle f, T_i g_k \rangle = \int_{\mathcal{M}} T_i g_k(y) f(y) d\mu(y)$$

The analysis coefficients c cannot be computed in practice. However, if we assume that the node of the graph arise from a uniform sampling distribution, this integral can be approximated by the discrete sum of the known samples, i.e:

$$c[i,k] = \mathbf{A}_{\mathcal{M},g} f[i,k] \approx \sum_{j=1}^{N} T_i g_k(x_j) f(x_j) \tag{10}$$

which is simply the analysis operator over the graph samples. Moreover, for kernels g_k satisfying $g_k(d) = 0$, respectively $g_k(1) = 0$ (see Theorem 8), we have $T_i g_k(x_j) = \boldsymbol{T}_i g_k(j)$. In that case, it is completely equivalent to the traditional analysis operator on graphs.

Definition 4 (Synthesis operator)**.** *Given a graph filter-bank $\{g_k\}_{k \in 1 \dots K}$ and a set of coefficient $c \in \mathbb{R}^{nK}$, the "out of sample" synthesis operator is defined as $\mathbf{S}_{\mathcal{M},g} : \mathbb{R}^{nK} \to \mathcal{M}$ is defined as*

$$\mathcal{S}_{\mathcal{M},g} c(y) = \sum_{k=1}^{K} \sum_{i=1}^{n} c_{i,k} T_i g_k(y) \tag{11}$$

The synthesis operator is the main object allows us to extend a signal out of the graph.

Span of the space of the synthesis operator Let us consider the function $s \in \mathcal{M}$ that is the result of a synthesis operation of the coefficients c.

$$\begin{aligned} s(y) &= \sum_{k=1}^{K}\sum_{i=1}^{n} c_{ik} T_i g_k(y) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{N} c_{ik} \sum_{\ell=1}^{n} g_k(\lambda_\ell)\boldsymbol{v}_\ell[i]u_\ell(x) \\ &= \sum_{\ell=1}^{n} u_\ell(x) \sum_{k=1}^{K} g_k(\lambda_\ell) \sum_{i=1}^{n} c_{ik}\boldsymbol{v}_\ell[i] \\ &= \sum_{\ell=1}^{n} \hat{\boldsymbol{c}}[\ell]u_\ell(x), \end{aligned}$$

where $\hat{\boldsymbol{s}} = \hat{\boldsymbol{c}}$ only if $g(d_y) = 0$. To be proved!

This show that the out of sample extension spans a space of size z given by the point-cloud Laplacian eigenvectors u_ℓ.

2.2 Main theorems

Using previous results, we obtain a convergence warranty on the defined atoms, i.e: a scaled version of $T_i g(y)$ converges toward $\mathcal{T}_{\mathcal{M}x_i} g(y)$.

Theorem 7 (Atoms convergence)**.** *Let us consider a continuously derivable band-limited kernel $g : \mathbb{R}_+ \to \mathbb{R}$, i.e: $\exists c < \infty | \forall \ell > cg(\lambda_e ll) = 0$. Let us defined a scaled version of g as $g^t(x) = g\left(\frac{\mathrm{vol}(\mathcal{M})}{t_n(4\pi t_n)^{\frac{k}{2}}}x\right)$. Then, using $k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{4t}}$, for t sufficiently small, $T^t_{i,n} g^t(y)$ exists and there exists a sequence $t(n) \to 0$, such that the*

$$\lim_{n\to\infty} T^t_{i,n} g^t(y) = \mathcal{T}_{\mathcal{M}x_i} g(y)$$

where the conergence is in probability

Proof. For all $\ell \leq c$, using Theorem 2, we know that $\lambda^t_{\ell,n}$ exists for a t sufficiently small. As a result, $T^t_{i,n} g^t(y)$ is well defined for this t.

Convergence in probability is preserved when applicated to continuous function. We can thus apply Theorem 2 and get the desired results:

$$\begin{aligned} \lim_{n\to\infty} \left(T_i g^t\right)^t_n (y) &= \lim_{n\to\infty} \sum_{\ell=0}^{n-1} g\left(\frac{\mathrm{vol}(\mathcal{M})}{t_n(4\pi t_n)^{\frac{k}{2}}}\lambda^{t_n}_{\ell,n}\right) u^{t_n}_{\ell,n}(x_i) u^{t_n}_{\ell,n}(y) \\ &= \sum_{\ell=0}^{\infty} g(\lambda_\ell) u_\ell(x_i) u_\ell(y) \\ &= \mathcal{T}_{\mathcal{M}x_i} g(y) \end{aligned}$$

□

Hidden behind the simplicity of the previous demonstration, there is a limit on the kernel. In fact Theorem 3 is used in the demonstration of Theorem 2. The eigenvectors $u_\ell(y)$ of L are defined only if $\lambda \notin \mathrm{rg}(\boldsymbol{d})$. The demonstration of Theorem 2 is based on the fact that $\mathrm{rg}(\boldsymbol{d}) \to \infty$ as $t \to 0$ (Proposition 5.1, 2).

For a fixed number of point n, Definitions 6 and 2 are defined for any continuously differentiable kernel g. However it is limited by the existence of the the eigenvectors u_ℓ of L, which are well defined for $\lambda \notin \mathrm{rg}(\boldsymbol{d})$. The following theorem quantify this effect:

Theorem 8 (Continuity of the atoms)**.** *Let $g : \mathbb{R}^+ \to \mathbb{R}$ be a derivable function and x_j a point of the graph with degree $\boldsymbol{d}[j] \neq \lambda_\ell, \forall \ell$, then the following holds:*

$$T_i g(x_j) = h_j(\boldsymbol{L})\boldsymbol{W}_j = \begin{cases} \boldsymbol{T}_i g[j] & \text{if } i \neq j \\ \boldsymbol{T}_i g[j] - g(\boldsymbol{d}[j]) & \text{if } i = j \end{cases}$$

where $h_j(x) = -\frac{g(\boldsymbol{d}[j]) - g(x)}{\boldsymbol{d}[j] - x}$. For the normalized case, we have:

$$T_i' g(x_j) = h_j(\boldsymbol{L})\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{W}_j \boldsymbol{d}[j]^{-\frac{1}{2}} = \begin{cases} \boldsymbol{T}_i g[j] & \text{if } i \neq j \\ \boldsymbol{T}_i g[j] - g(1) & \text{if } i = j \end{cases}$$

where $h_j(x) = -\frac{g(1) - g(x)}{1 - x}$.

The proof is given in Appendix C.

Theorem 8 warranty that low frequency based atoms (i.e: $g(x) = 0, x < d \Rightarrow g(\boldsymbol{d}[j]) = 0$) are continuous within the graph localization operator $\boldsymbol{T}_i g$. However, when they present some high frequency content (i.e: $g(\boldsymbol{d}[j]) \neq 0$), a single discontinuity appears at the node i.

Example 4 (Ring graph)**.** *Let us consider a ring graph with uniform weight of 0.5 and self loops of weight 1. The degree of every node is $d = 2$ and the maximum eigenvalue is $\boldsymbol{\lambda}_{\max} = 2$. As a result, the maximum eigenvalue is 2. In this case, the sampling limit would be $\boldsymbol{\lambda}_{\max}$ which the actually the Shannon bound (because the graph spectrum in folded in the positive frequencies).*

Example 5. *Let us consider a wavelet fitlerbank. In figure 4 TBD*

2.3 Relation with Nyström interpolation method

The following assertions are inspired from the paper [7], which is itself based on [12, 10, 1].

The Nyström technique allows us to find numerical approximation of an eigenfunction problem defined on a continuous space. For simplicity, we present the one dimensional case. Give a kernel $k[a, b]^2 \to \mathbb{R}$, the problem is defined as the eigenfunctions $\phi(x)$ and eigenvalues λ such that:

$$\int_a^b k(x, y)\phi(y)\mathrm{d}y = \lambda\phi(x) \tag{12}$$

This integral can be approximated by a set of evenly spaced points $x_1, \ldots, x_n$ to become:

$$\frac{b - a}{n}\sum_{i=1}^{n} k(x, x_i)\tilde{\phi}(x_i) = \lambda\tilde{\phi}(x), \tag{13}$$

where $\tilde{\phi}$ is an approximation of the true eigenfunction $\phi(x)$. Equation (13) can be solved using only the subset of points x_i:

$$\frac{b - a}{n}\sum_{i=1}^{n} k(x_j, x_i)\tilde{\phi}(x_i) = \tilde{\lambda}\tilde{\phi}(x_j), \qquad \forall j \in \{1, \ldots n\},$$

or, more conveniently, in a matrix form:

$$cK\tilde{\Phi} = \tilde{\Phi}\tilde{\Lambda},$$

where $c = \frac{b-a}{n}$ and K_{ij} is the kernel k evaluated at x_i, x_j. Using traditional techniques, one can solve this eigenfunction problem to obtain $\tilde{\lambda}_\ell$ and $\tilde{\phi}_\ell[j], \forall j \in \{1, \ldots n\}$ and $\forall \ell \in \{0, \ldots n - 1\}$. The final step consists of inserting those variable inside (13) to obtain the first n eigenvectors interpolation:

$$\hat{\phi}_\ell(y) = \frac{1}{\tilde{\lambda}_\ell}\sum_{i=1}^{n} k(y, x_i)\hat{\phi}_\ell(x_i) \tag{14}$$

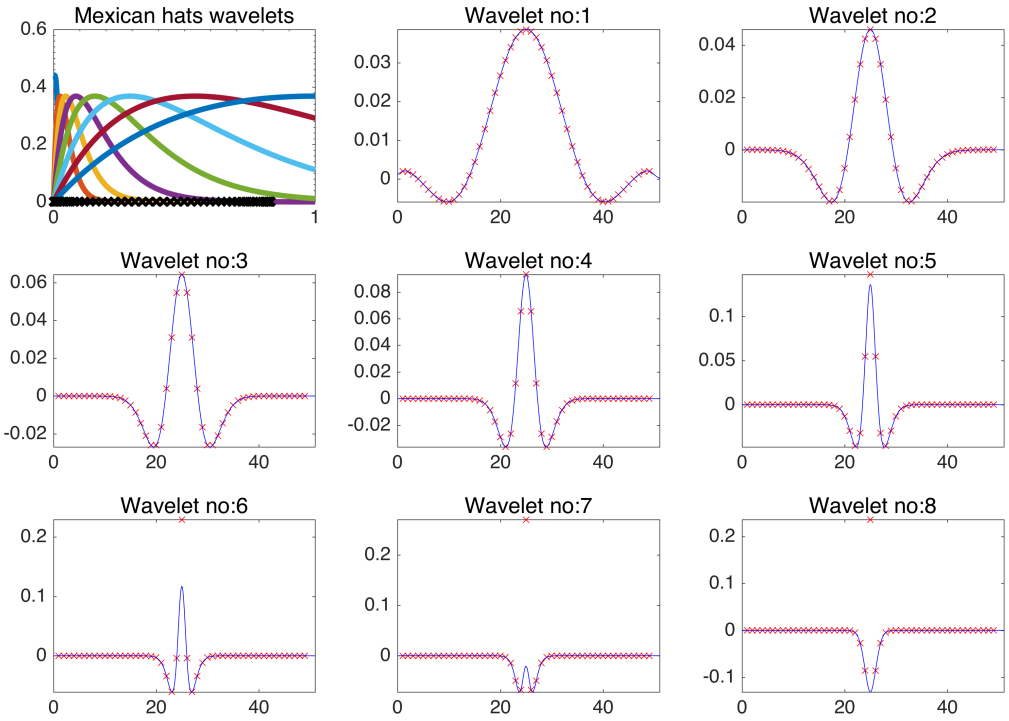

Figure 4: Extented wavelet atoms using the normalized definition (**??**). The difference between the extension and the central vertex is exactly the value of the kernel at 1.

Nyström interpolation has been used in [5] to extend out of the samples lle, isomap, mds, eigenmaps, and spectral clustering. The method used inside [5] is a simple generalization of what is presented above taking care of the irregular sampling. The extended eigenvectors used are given in (14).

Comparison with the point cloud Laplacian eigenvectors U. We observe that there is a strong similarity between the Nyström interpolated eigenvectors (14) and the point cloud Laplacian eigenvectors (7) and (8). The latter can be considered as extension of the Nyström for an irregular sampling.

In the normalized case, we have

$$\boldsymbol{I} - \boldsymbol{K}' = \boldsymbol{L}' = \boldsymbol{U}'\boldsymbol{\Lambda}'\boldsymbol{U}'^{*} = \boldsymbol{U}'(\boldsymbol{I} - \boldsymbol{\Omega})\boldsymbol{U}'^{*},$$

where $\boldsymbol{\Omega}$ are the eigenvalues of the kernel matrix $\boldsymbol{K}'$. Thus we have $\boldsymbol{\Omega} = \boldsymbol{I} - \boldsymbol{\Lambda}$. From this, we observe that in the normalized case, the expression are equivalent.

3 Extending a signal residing on a graph

Using the theory presented above, we propose a method to extend graph signal out of the graphs. It allow us to extend the result of processing data to new sample with a minimum number of computation. In this section, we develop a method that can be used

as a "black box" to extend typical graph machine learning problems such as de-noising, classification or regression. The utility of this contributions does not limit to this method. For example, we will show later how spectral clustering can be accelerated.

The method. The method can be summarized as follow. First compute the graph Fourier transform of the signal. Then weights the extended Fourier eigenvectors with the resulting coefficients.

Let consider a signal with graph sample denoted $\boldsymbol{s}$. The Fourier transform becomes

$$\hat{\boldsymbol{s}} = \boldsymbol{U}^* s$$

Let us consider the line vector $\boldsymbol{w}_y[i] = k(y, x_i)$. Equation (7) can be rewritten as:

$$u_\ell(y) = [\boldsymbol{w}_y^T \boldsymbol{U}(d\boldsymbol{I} - \boldsymbol{\Lambda})^{-1}]_\ell,$$

The out of sample extension is defined as a weighed sum of all Fourier mode:

$$s(y) = \sum_{\ell=1}^{n} \hat{\boldsymbol{s}}[\ell] u_y[\ell] = \boldsymbol{w}_y^T \boldsymbol{U}(d\boldsymbol{I} - \boldsymbol{\Lambda})^{-1} \boldsymbol{U}^* s = \boldsymbol{w}_y^T (d\boldsymbol{I} - \boldsymbol{L})^{-1} s$$

Link with RKHS.

$$s = K\alpha$$

$$\arg\min_\alpha \|K\alpha - y\|^2 + \gamma \alpha^T K \alpha$$

The solution satisfies

$$K^2\alpha - Ky + K\alpha = 0$$

$$\begin{aligned} \alpha &= (K^2 + \gamma K)^{-1} K y \\ &= (K + \gamma I)^{-1} y \end{aligned}$$

Remember that for the nomalized case

$$L = I - K$$

As a result, we find

$$\begin{aligned} \alpha &= (K + \gamma I)^{-1} y \\ &= (I(1+\gamma) - L)^{-1} y \end{aligned}$$

This expression shows that the graph out of sample extension is equivalent to solve

$$\arg\min_\alpha \|K\alpha - y\|^2.$$

Note about how to present the result.

- We should not be against the RKHS but complementary to it.
- Let us suppose that the kernel instead of being K would be someting else.
  - K is spsd
  - $L = D - K$ is spsd to

  - Let choose $h : \mathbb{R}^+ \to \mathbb{R}^+$, then $h(L)$ is spsd, which implies that $h(L)$ is a valid kernel!
- We have a way to extend $h(L)$
- Questions:
  - Is the extension of $h(L)$ a RKHS? No
  - Under a stationarity assumption, how close are the extended $h(L)$ from the original atoms (with respect of the number of samples)? We can probably answer this question using Smale paper.
- We need stationarity for manifold!
  - Stationary signal on a manifold
  - Approximate everything with a graph and push it pack to the manifold.

Graph efficient construction. The theory presented above is constructed with a kernel k that is assumed to be grater than 0 for every pair of point of $\mathcal{M}$. However, this assumption implies to work with fully connected graphs. Providing that the graph remain connected, the theory remains compatible with ϵ-radius graph. In this case, the kernel is set to 0 for every pair of nodes with distance greater than ϵ. Problematically, finding all sample in a ϵ neighborhood is still an expensive operation. We have tested our method with only the k nearest neighbors. Also not all the theoretical results hold in this case, practical results seem to behave as expected.

Out of sample extension via Chebyshev polynomials TBD

Complexity. TBD

4 Illustrative applications

4.1 Interpolation

4.2 Extended spectral clustering

4.3 Semi-supervised learning

4.4 Manifold estimation

5 Conclusion

Acknowledgement

This has been supported by the SNF research project *Towards Signal Processing on Graphs*, grant number: 2000_21/154350/1.

References

[1] C. T. Baker and C. Baker. *The numerical treatment of integral equations*, volume 13. Clarendon press Oxford, 1977.

[2] M. Belkin and P. Niyogi. Towards a theoretical foundation for laplacian-based manifold methods. In *Learning theory*, pages 486–500. Springer, 2005.

[3] M. Belkin and P. Niyogi. Convergence of laplacian eigenmaps. *Advances in Neural Information Processing Systems*, 19:129, 2007.

[4] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research*, 7:2399–2434, 2006.

[5] Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. *Advances in neural information processing systems*, 16:177–184, 2004.

[6] F. R. Chung. *Spectral graph theory*, volume 92. AMS Bookstore, 1997.

[7] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the nystrom method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):214–225, 2004.

[8] D. Geller and A. Mayeli. Continuous wavelets on compact manifolds. *Mathematische Zeitschrift*, 262(4):895–927, 2009.

[9] D. K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.

[10] E. J. Nyström. *Über die praktische Auflösung von linearen Integralgleichungen mit Anwendungen auf Randwertaufgaben der Potentialtheorie*. Akademische Buchhandlung, 1929.

[11] N. Perraudin, R. Benjamin, S. David I, and P. Vandergheynst. Global and local uncertainty principles for signals on graphs. *arXiv preprint arXiv:1603.03030*, 2016.

[12] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*, volume 2. Cambridge university press Cambridge, 1996.

[13] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *Signal Processing Magazine, IEEE*, 30(3):83–98, 2013.

[14] D. I. Shuman, B. Ricaud, and P. Vandergheynst. Vertex-frequency analysis on graphs. arXiv:1307.5708, 2013.

[15] U. Von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, pages 555–586, 2008.

A Usual definitions

Definition 5 (Range of a function)**.** *The range of a continuous function $f : X \to Y$ (where X is a compact set) is the smallest closet set of Y where that contains $f(X)$. It is defined by:*

$$\operatorname{rg}(f) = \left[\inf_x f(x), \sup_x f(x)\right].$$

For example the range of the degree is simply $\operatorname{rg}(d) = [d_{\min}, d_{\max}]$. We use *compact* to signifies closed and bounded. We denote the spectrum of an operator T by $\sigma(T)$.

Definition 6 (Discrete spectrum)**.** *The* discrete spectrum *of an operator T denoted $\sigma_d(T)$ is the set of all isolated eigenvalues with finite multiplicity.*

The *essential spectrum* is the rest of the eigenvalues: $\sigma_{\text{ess}}(T) := \sigma(T) \backslash \sigma_d(T)$. The essential spectrum is always closed and the discrete spectrum can only have accumulation points on the boundary to the essential spectrum. The essential spectrum is invariant under compact perturbations. That is, if K is a compact self-adjoint operator, then we have $\sigma_{\text{ess}}(T) = \sigma_{\text{ess}}(T + K)$. Weyl's define the spectrum and the essential spectrum in the following way. λ is in the spectrum of T if and only if there exists a sequence ϕ_k such that $\|\phi_k\| = 1$ and

$$\lim_{k\to\infty} \|T\phi_k - \lambda\phi_k\| = 0.$$

Furthermore, λ is in the essential spectrum if there is a sequence satisfying this condition, but such that it contains no convergent subsequence. If we suppose that λ is in the discrete spectrum, all sequences have to jump discretely from one eigenvector repeat at some point ϕ_λ (this is the case if, for example $\{\phi_k\}$ is an orthonormal sequence); We call such a sequence a singular sequence.

B Spectral decomposition of the point-cloud Laplacian

TO DO: proofs for the normalized case Although, the point-cloud Laplacian is not symmetric, it possesses many very convenient spectral properties. First, it has the same eigenvalues as $\boldsymbol{L}$ and an additional one: d.

Lemma 1. *Eigenvalues of the extended graph Laplacian*

$$\Lambda(L) = \boldsymbol{\Lambda}(\boldsymbol{L}) \cup \{d\}$$

Proof. To compute the eigenvalue of the new operator, we can find the root of the determinant of the following matrix:

$$A = \begin{pmatrix} -\lambda I & 0 \\ -w & d-\lambda \end{pmatrix}$$

Since A is a block matrix, we have:

$$\det(A) = \det(\boldsymbol{L}) \cdot (d - \lambda)$$

As a result, the eigenvalues of L are the ones from $\boldsymbol{L}$ and d. □

Second, the eigenvectors of L include the graph Fourier basis.

Lemma 2. *Eigenvectors of the extended graph laplacian Suppose $d \neq \lambda$, $\forall\lambda \in \boldsymbol{\Lambda}(\boldsymbol{L})$. Then the eigenvectors of L are:*

$$\boldsymbol{U} = \begin{pmatrix} U & 0 \\ c & 1 \end{pmatrix}$$

with $c_i = \frac{\sum_{j=1}^n w_j U_{ij}}{\lambda_i - d} = \frac{(U^* w)_i}{\lambda_i - d}$.

Proof. We simply solve the eigenvalue system of equations

$$\begin{pmatrix} U & 0 \\ -wU + dc & d \end{pmatrix} = \begin{pmatrix} & 0 \\ -w & d \end{pmatrix} \begin{pmatrix} U & 0 \\ c & 1 \end{pmatrix} = \begin{pmatrix} U & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} \Lambda & 0 \\ 0 & d \end{pmatrix} = \begin{pmatrix} U\Lambda & 0 \\ c\Lambda & d \end{pmatrix}$$

The only non trivial part is $-wU + dc = c\Lambda$. It implies that

$$c = wU(dI - \Lambda)^{-1}.$$

□

From now on, we will denote $\boldsymbol{U}, \boldsymbol{\Lambda}$ as the eigenvectors, eigenvalues matrices of $\boldsymbol{L}$. The matrix $\boldsymbol{U}$ is invertible.

Lemma 3. *The inverse of $\boldsymbol{U}$ is given by*

$$\boldsymbol{U}^{-1} = \begin{pmatrix} U^* & 0 \\ -cU^* & 1 \end{pmatrix}$$

Proof. $\boldsymbol{U}$ is always invertible since it has the same determinant than U and U is invertible with inverse U^*.

$$\begin{pmatrix} U^* & 0 \\ -cU^* & 1 \end{pmatrix} \begin{pmatrix} U & 0 \\ c & 1 \end{pmatrix} = \begin{pmatrix} U^*U & 0 \\ -cU^*U + c & 1 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & 1 \end{pmatrix}$$

□

As a result, the matrix $\boldsymbol{L}$ can be decomposed as

$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1} = \begin{pmatrix} U & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} \Lambda & 0 \\ 0 & d \end{pmatrix} \begin{pmatrix} U^* & 0 \\ -cU^* & 1 \end{pmatrix}$$

C Proof of Theorem 8

Proof. Let us first separate the filter h_j in $h_{j1}(x) = -\frac{g(d_j)}{d_j - x}$ and $h_{j2}(x) = \frac{g(x)}{d_j - x}$. From Lemma 4, we find

$$h_{j1}(\boldsymbol{L})\boldsymbol{W}_j = -g(d_j)\delta_j$$

From Lemma 5, we have directly that

$$h_{j2}(\boldsymbol{L})\boldsymbol{W}_j = \boldsymbol{T}_j g$$

When $i = j$, the we find naturally

$$\left(h_j(\boldsymbol{L})\boldsymbol{W}_j\right)(i) = \left(h_{j1}(\boldsymbol{L})\boldsymbol{W}_j\right)(i) + \left(h_{j2}(\boldsymbol{L})\boldsymbol{W}_j\right)(i) = -g(d_j) + \boldsymbol{T}_i g(i).$$

In the other case $h_{j1}(\boldsymbol{L})\boldsymbol{W}_j = 0$ and

$$\left(h_j(\boldsymbol{L})\boldsymbol{W}_j\right)(i) = \boldsymbol{T}_j g(i) = \boldsymbol{T}_i g(j)$$

□

Lemma 4. *Let select the node i such that $\boldsymbol{d}_i \neq \lambda_\ell, \forall \ell$, then the kernel $k(x) = \frac{1}{x - d_i}$ satisfies*

$$k(\boldsymbol{L})\boldsymbol{W}_i = \delta_i$$

Proof. Because d_i is not equal to an eigenvalue, $k(\boldsymbol{L})$ is invertible and we have

$$k(\boldsymbol{L})^{-1}\delta_i = (\boldsymbol{L} - Id_i)\delta_i = \boldsymbol{L}_i - I_i d_i = \boldsymbol{W}_i$$

□

Lemma 5. *Let select the node i such that $d_i \neq \lambda_\ell, \forall \ell$, then the kernel $h_i(x) = -\frac{g(x)}{x-d_i}$ satisfies*

$$\boldsymbol{T}_i g = h_i(\boldsymbol{L})\boldsymbol{W}_i$$

Proof. Since $d_i \neq \lambda_\ell$, we know that the kernel $a_i(x) = \frac{1}{x-d_i}$ is well defined and invertible. As a result, we need to show that $a^{-1}(\boldsymbol{L})\boldsymbol{T}_i g = a^{-1}(\boldsymbol{L})h_i(\boldsymbol{L})\boldsymbol{W}_i$, i.e:

$$\boldsymbol{T}_i t_i = -g(\boldsymbol{L})\boldsymbol{W}_i \tag{15}$$

where $t_i(x) = g(x)(x - d_i)$. Let us expand this last expression spectrally. On the left side, we find

$$\begin{aligned} \boldsymbol{T}_i t_i &= U\left(g(\Lambda)(\Lambda - d_i I)\right) U^* \delta_i \\ &= U g(\Lambda)\Lambda U^* \delta_i - d_i U g(\Lambda) U^* \delta_i, \end{aligned}$$

and on the right side we have

$$\begin{aligned} -g(\boldsymbol{L})\boldsymbol{W}_i &= -U\left(g(\Lambda)\right) U^* \boldsymbol{W}_i \\ &= U\left(g(\Lambda)\right) U^* (\boldsymbol{L} - D)\delta_i \\ &= U\left(g(\Lambda)\right) U^* U \Lambda U^* \delta_i - U\left(g(\Lambda)\right) U^* D \delta_i \\ &= U g(\Lambda)\Lambda U^* \delta_i - d_i U g(\Lambda) U^* \delta_i. \end{aligned}$$

Identifying term by term both side of (15) concludes the proof. □