Proceedings of the World Congress on Engineering 2010 Vol I
WCE 2010, June 30 - July 2, 2010, London, U.K.

On An Interpretation Of Spectral Clustering Via Heat Equation And Finite Elements Theory

Sandrine Mouysset, Joseph Noailles and Daniel Ruiz *

Abstract— **Spectral clustering methods use eigenvectors of a matrix, called Gaussian affinity matrix, in order to define a low-dimensional space in which data points can be clustered. This matrix is widely used and depends on a free parameter σ. It is usually interpreted as some discretization of the Heat Equation Green kernel. Combining tools from Partial Differential Equations and Finite Elements theory, we propose an interpretation of this spectral method which offers an alternative point of view on how spectral clustering works. This approach develops some particular geometrical properties inherent to eigenfunctions of some specific partial differential equation problem. We analyze experimentally how this geometrical property is recovered in the eigenvectors of the affinity matrix and we also study the influence of the parameter σ.**

Keywords: Clustering, Machine Learning, Spatial Data Mining

1 Introduction

Clustering aims to partition a data set by grouping similar elements into subsets. Two general main issues concern, on the one hand, the choice of a similarity criterion and, on the other hand, how to separate clusters the one from the other. Spectral methods, and in particular the spectral clustering algorithm introduced by Ng-Jordan-Weiss (NJW) [8], are useful when considering non-convex shaped subsets of points. Spectral clustering (SC) consists in defining a low-dimensional data space by selecting particular eigenvectors of a matrix called normalized affinity matrix in which data points are clustered. In this clustering method, two main problems arise : first, what sense can we give to the notion of cluster when considering finite discrete data sets and then, how can we link these clusters to some spectral elements (eigenvalues/eigenvectors extracted in SC).
Spectral Clustering can be analyzed using graph theory and some normalized cut criterion leading to the solution of some trace-maximum problems (see, for instance, [8], [9]). As the aim is to group data points in some sense of closeness, Spectral Clustering can also be interpreted by exploiting the neighbourhood property in the discretization of the Laplace-Beltrami operator and heat kernel on manifolds [2] : in the algorithm, neighbourhood property is represented by the step of adjacency graph. Moreover, consistency of the method was investigated by considering the asymptotic behaviour of clustering when samples become very large [11, 10]. Some properties under standard assumptions for the normalized Laplacian matrices were proved, including convergence of the first eigenvectors to eigenfunctions of some limit operator [11, 7]. To the best of our knowledge, the main results establish this relation for huge numbers of points. However, from a numerical point of view, SC still works for smaller data sets. So, in this paper, we try to give explanations that may address a given data sample.
A second problem appears when considering the most widely used affinity matrix in spectral clustering techniques, as in the NJW-algorithm. It is based on the Gaussian measure between points which depends on a free scale parameter σ which has to be properly defined. Many investigations on this parameter were led and several definitions were suggested either heuristically [12, 3], or with physical considerations [6], or from geometrical point of views [5]. The difficulty to fix this choice seems to be tightly connected to the lack of some clustering property explaining how the grouping in this low-dimensional space defines correctly the partitioning in the original data. We will show how this free parameter affects clustering results.
In this paper, as spectral elements used in SC do not give explicitly this topological criteria for a discrete data set, we are drawing back to some continuous formulation wherein clusters will appear as disjoint subsets. So eigenvectors of the affinity matrix A will be interpreted as eigenfunctions of some operator. This drawback is performed (section 3) using Finite Elements (FE). Indeed, with FE whose nodes correspond to data points, representation of any L^2-function is given by its nodal value. So we expect to interpret both A and its eigenvectors as a representation of respectively a L^2 operator and L^2 functions on some bounded regular Ω.
Now, the commonly used operator whose Finite Elements representation matches with A is the kernel of the Heat equation on unbounded space (noted K_H). As its spectrum is essential, we cannot interpret eigenvectors of A as a representation of eigenfunctions of K_H. However, on a bounded domain $\mathcal{O}$, K_H is closed to K_D, the kernel of the Heat equation on a bounded domain Ω, for Ω including

*IRIT-ENSEEIHT, 2 rue Charles Camichel, University of Toulouse, France, 31017 BP 122
{sandrine.mouysset,joseph.noailles,daniel.ruiz}@enseeiht.fr

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

strictly $\mathcal{O}$ (section 2.2). Now, operator S_D (convolution by K_D) has eigenfunctions (v_{n_i}) in $H_0^1(\Omega)$. From this, we show (proposition 1) that operator S_H (convolution by K_H) admits these v_{n_i} as near eigenfunctions plus a residue (noted η). At last, using Finite Elements approximation, we show (proposition 2) that eigenvectors of A are a representation of these eigenfunctions plus a residue (noted ψ).

To summarize, the main result of this paper is that for a fixed data set of points, the eigenvectors of A are the representation of functions which support is included in only one connected component at once. The accuracy of this representation is shown to depend for a fixed density of points, on the affinity parameter t. This result is illustrated with small fixed data points (section 4).

1.1 Spectral clustering Algorithm

Let us first give some notations and recall the NJW algorithm for a n points data set in a p-dimensional euclidean space. Assume that the number of targeted clusters k is known. The algorithm contains few steps which are described as follows:

Algorithm 1 Spectral Clustering Algorithm

Input : data set, number of clusters k

1. Form the affinity matrix $A \in \mathbb{R}^{n\times n}$ defined by:
$$A_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) & \text{if } i \neq j, \\ 0 \text{ otherwise,} \end{cases} \quad (1)$$
2. Construct the normalized matrix: $L = D^{-1/2}AD^{-1/2}$ with $D_{i,i} = \sum_{j=1}^{n} A_{ij}$,
3. Assemble the matrix $X = [X_1 X_2 .. X_k] \in \mathbb{R}^{n\times k}$ by stacking those eigenvectors associated with the k largest eigenvalues of L,
4. Form the matrix Y by normalizing each row in the $n \times k$ matrix X,
5. Treat each row of Y as a point in $\mathbb{R}^k$, and group them in k clusters via the *K-means* method,
6. Assign the original point x_i to cluster j when row i of matrix Y belongs to cluster j.

In this paper, we won't take into account the normalization of the affinity matrix described in step 2. But we focus on the two main steps of this algorithm: assembling the Gaussian affinity matrix and extracting its eigenvectors to create the low-dimensional space.

2 Clustering property of Heat kernel

As recalled in introduction, Spectral Clustering consists in selecting particular eigenvectors of the normalized affinity matrix defined by (1). Gaussian affinity coefficients A_{ij} are usually interpreted as a representation of the Heat kernel evaluated at data points x_i and x_j.

Let $K_H(t,x,y) = (4\pi t)^{-\frac{p}{2}} \exp\left(-\frac{\|x-y\|^2}{4t}\right)$ be the Heat kernel in $\mathbb{R}_+^* \times \mathbb{R}^p \times \mathbb{R}^p$. Note that the Gaussian affinity between two distinct data points x_i and x_j is defined by $A_{ij} = (2\pi\sigma^2)^{-\frac{p}{2}}\ K_H(\sigma^2/2, x_i, x_j)$. Now, consider the initial value problem in $L^2(\mathbb{R}^p)$, for $f \in L^2(\mathbb{R}^p)$:

$$(\mathcal{P}_{\mathbb{R}^p}) \begin{cases} \partial_t u - \Delta u = 0 \ \text{ for } \ (x,t) \in \mathbb{R}^p \times \mathbb{R}^+ \backslash \{0\}, \\ u(x,0) = f \ \text{ for } x \in \mathbb{R}^p, \end{cases}$$

and introduce the solution operator in $L^2(\mathbb{R}^p)$ of $(\mathcal{P}_{\mathbb{R}^p})$:

$$(S_H(t)f)(x) = \int_{\mathbb{R}^p} K_H(t,x,y) f(y) dy, \ x \in \mathbb{R}^p. \quad (2)$$

Now if $f(y) \approx \delta_{x_j}(y)$, where δ_{x_j} denotes the Dirac function on x_j, one can observe that:

$$\left(S_H(\frac{\sigma^2}{2})f\right)(x_i) \approx K_H\left(\frac{\sigma^2}{2}, x_i - x_j\right) = \left(2\pi\sigma^2\right)^{\frac{p}{2}} (A + I_n)_{ij}.$$

Thus, the spectral properties of matrix A used in the Spectral clustering algorithm seems to be related to those of operator $S_H(t)$. We propose to analyze this in details in the following.

2.1 Heat equation with Dirichlet boundary conditions

To simplify, we shall fix in the following the number of clusters k to 2, but the discussion can be extended to any arbitrary number k. Consider then a bounded open set Ω in $\mathbb{R}^p$ made up of two disjoint connected open subsets Ω_1 and Ω_2. Assume that the boundary $\partial\Omega = \partial\Omega_1 \cup \partial\Omega_2$ with $\partial\Omega_1 \cap \partial\Omega_2 = \emptyset$ is regular enough so that the trace operator is well-posed on $\partial\Omega$ and the spectrum decomposition of Laplacian operator is well-defined. For instance, $\partial\Omega$ is C^1 on both $\partial\Omega_1$ and $\partial\Omega_2$. According to [4], determining the eigenvalues $(\lambda_n(\Omega))_{n>0}$ and associated eigenfunctions $(v_n)_{n>0}$ of the Dirichlet Laplacian on Ω:

$$(\mathcal{P}_\Omega^L) \begin{cases} \Delta v_n = \lambda v_n \text{ in } \Omega, \\ v_n = 0 \text{ on } \partial\Omega, \end{cases}$$

is equivalent to define a function of the eigenvalues $(\lambda_{n_1}(\Omega_1))_{n_1>0}$ and $(\lambda_{n_2}(\Omega_2))_{n_2>0}$ and of the eigenfunctions of the Dirichlet Laplacian on Ω_1 and Ω_2 respectively. In other words, if $v_n \in H_0^1(\Omega)$, then $\Delta v = \lambda v$ on Ω if and only if for $i \in \{1,2\}$, $v_i = v|_{\Omega_i} \in H_0^1(\Omega_i)$ satisfies $\Delta v_{n_i} = \lambda v_{n_i}$ on Ω_i. Therefore, $\{\lambda_n(\Omega)\}_{n>0} =$

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$\{\lambda_{n_1}(\Omega_1)\}_{n_1>0} \cup \{\lambda_{n_2}(\Omega_2)\}_{n_2>0}$. Additionally, if we consider v_{n_i} the solution on Ω_i of $\mathcal{P}^L_{\Omega_i}$, an eigenfunction of the Laplacian operator Δ associated to λ_{n_i}, and if we extend the support of $v_{n_i} \in \Omega_i$ to the whole set Ω by setting $v_{n_i} = 0$ in $\Omega \backslash \Omega_i$, we also get an eigenfunction of the Laplacian Δ in $H^1_0(\Omega)$ associated to the same eigenvalue λ_{n_i}. Note that since Ω_1 and Ω_2 are disjoint, the extension by 0 to the whole set Ω of any function in $H^1_0(\Omega_i)$ is a function in $H^1_0(\Omega)$. Consequently, the union of the two sets of eigenfunctions $\{(v_{n_i})_{n_i>0},\ i \in \{1,2\}\}$, is an Hilbert basis of $H^1_0(\Omega)$.

Now let $f \in L^2(\Omega)$ and consider the heat problem with Dirichlet conditions on the bounded set $\Omega = \Omega_1 \cup \Omega_2$:

$$(\mathcal{P}^D_\Omega) \begin{cases} \partial_t u - \Delta u = 0 \text{ in } \mathbb{R}^+ \times \Omega, \\ u = 0, \text{ on } \mathbb{R}^+ \times \partial\Omega, \\ u(t=0) = f, \text{ in } \Omega. \end{cases}$$

Denote by K^Ω_D the Green's kernel of $(\mathcal{P}^D_\Omega)$. The solution operator in $H^2(\Omega) \cap H^1_0(\Omega))$ associated to this problem is defined, for $f \in L^2(\Omega)$, by:

$$S^\Omega_D(t)f(x) = \int_\Omega K_D(t,x,y)f(y)dy,\ x \in \mathbb{R}^p.$$

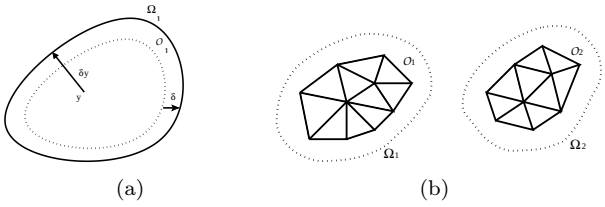
Eigenfunctions of $S^\Omega_D(t)$ and those of the Laplacian operator Δ are the same. Indeed, if v_{n_i} is an eigenfunction of $(\mathcal{P}^L_\Omega)$ associated to an eigenvalue λ_{n_i} then v_{n_i} is also eigenfunction of $S_D(t)$ with eigenvalue $e^{t\lambda_{n_i}}$ in the sense that $S^\Omega_D(t)v_{n_i}(x) = e^{t\lambda_{n_i}}v_{n_i}(x)$. Thus the geometrical property of the eigenfunction is preserved, *i.e* the support of eigenfunctions can be defined in only one connected component Ω_1 or Ω_2.

2.2 A clustering property from eigenfunctions of Heat equation with Dirichlet boundary conditions

We consider now the solution operator $S_H(t)$ of $(\mathcal{P}_{\mathbb{R}^p})$. As opposed to the previous case, we can not deduce similar properties as with the operator solution of $(\mathcal{P}^D_\Omega)$ because the spectrum of $S_H(t)$ is essential and eigenfunctions are not localized in $\mathbb{R}^p$ without boundary conditions. However, we are going to compare, for fixed values of t, $S_H(t)$ and $S^\Omega_D(t)$ and derive some asymptotic properties on $S_H(t)$ that approximate the previously analyzed properties of $S^\Omega_D(t)$.

For $\varepsilon > 0$, let introduce an open subset $\mathcal{O}$ which approximates from the interior the open set Ω, in the sense that $\mathcal{O} \subset \bar{\mathcal{O}} \subset \Omega$ and $\text{Volume}(\Omega \backslash \mathcal{O}) \leq \varepsilon$. Note $\delta(y)$ the distance of a point $y \in \mathcal{O}$ to $\partial\Omega$, the boundary of Ω, and $\delta = \inf_{y\in\mathcal{O}} \delta(y)$ the distance from $\mathcal{O}$ to Ω (see fig1). At last, $\mathcal{O}$ is supposed to be chosen such that $\delta > 0$.

As recalled in [1], for $t > 0$, x and y in Ω, $K_H(t,x,y)$ is a distribution that corresponds to the elementary solutions



(a) (b)

Figure 1: (a) Approximation of Ω_1 with an open set $\mathcal{O}_1$, (b) Mesh representation : $x_i \in \overset{\circ}{\Omega},\ \forall i \in \{1,..n\}$

of the heat equation $(\mathcal{P}_{\mathbb{R}^p})|_\Omega$, K_H is strictly positive on the boundary $\partial\Omega$ and, due to the maximum principle, K_D is bounded above by K_H, $\forall (t,x,y) \in \mathbb{R}_+ \times \Omega \times \Omega$:

$$0 < K_D(t,x,y) < K_H(t,x,y). \tag{3}$$

Furthermore, both Green's kernels K_H and $K^{\Omega_i}_D$ approximate each other on each open set $\mathcal{O}_i,\ i \in \{1,2\}$ and an estimation about their difference can be given by, $\forall (x,y) \in \mathcal{O}_i \times \mathcal{O}_i,\ \forall i \in \{1,2\}$:

$$0 < K_H(t,x,y) - K^{\Omega_i}_D(t,x,y) \leq \frac{1}{(4\pi t)^{\frac{p}{2}}} e^{-\frac{\delta(y)^2}{4t}}. \tag{4}$$

Now, for any $f \in L^2(\Theta)$, let us come back to the heat solution $S_H(t)f$ and restrict its support to the open set $\mathcal{O}_i$, for $i \in \{1,2\}$. For any open set Θ, the resulting restricted operator S_H to Θ, for $f \in L^2(\Theta)$ is:

$$S^\Theta_H(t)f(x) = \left(\int_\Theta K_H(t,x,y)f(y)dy \right) |_\Theta, x \in \Theta.$$

At last, define also $\mathcal{T}_i$, for $i \in \{1,2\}$ as the L^2-mapping from $L^2(\mathcal{O}_i)$ onto $L^2(\mathcal{O})$ which extends the support of any function $u \in L^2(\mathcal{O}_i)$ to the whole set Ω with:

$$\forall u \in L^2(\mathcal{O}_i),\ \mathcal{T}_i(u) = \begin{cases} u \text{ in } \mathcal{O}_i, \\ 0 \text{ in } \Omega \backslash \mathcal{O}_i \end{cases} \tag{5}$$

With these notations, we can then state the following proposition:

Proposition 1 *Let $i \in \{1,2\}$, $\widetilde{v_{n_i}} = \mathcal{T}_i(v_{n_i}|_{\mathcal{O}_i})$ where v_{n_i} is the eigenfunction in $H^2(\Omega_i) \cap H^1_0(\Omega_i)$ associated to $\exp^{-\lambda_{n_i}t}$ for operator $S^{\Omega_i}_D$. Then, for $t > 0$, we have the following result:*

$$S^\mathcal{O}_H(t)\widetilde{v_{n_i}} = \exp^{-\lambda_{n_i}t}\widetilde{v_{n_i}} + \eta(t,\widetilde{v_{n_i}}), \tag{6}$$

with $\|\eta(t,\widetilde{v_{n_i}})\|_{L^2(\mathcal{O})} \to 0$ when $t \to 0, \delta \to 0$.

Proof sketch. By triangular inequality, $\|\eta(t,\widetilde{v_{n_i}})\|_{L^2(\mathcal{O})} = \left\|S^\mathcal{O}_H(t)\widetilde{v_{n_i}} - \exp^{-\lambda_{n_i}t}\widetilde{v_{n_i}}\right\|_{L^2(\mathcal{O})}$ is bounded above by three terms:

$$\|\eta(t,\widetilde{v_{n_i}})\|_{L^2(\mathcal{O})} \leq \|\varsigma(t,v_{n_i}|_{\mathcal{O}_i})\|_{L^2(\mathcal{O})} + \dots$$

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$\dots \left\| S_D^{\Omega_i}(t)\mathcal{T}_i\left(v_{n_i}|_{\mathcal{O}_i} - v_{n_i}\right)\right\| + \left\| S_D^{\Omega_i}(t)v_{n_i} - e^{-\lambda_{n_i}t}\widetilde{v_{n_i}}\right\|_{L^2(\mathcal{O})}$ $t > 0$ by, $\forall y \in \mathcal{O},\ \forall\ 1 \leq j \leq n$:

On the right side, $\mathcal{T}_i\left(v_{n_i}|_{\mathcal{O}_i} - v_{n_i}\right)$ has its support included in $\Omega_i \backslash \mathcal{O}_i$ and *supp* $\Omega_i \backslash \mathcal{O}_i$ tends to $\emptyset$ when $\delta \to 0$. Moreover, operator $S_D^{\Omega_i}(t)$ tends to identity when $t \to 0$. So both two last terms in the right side of this inequality vanish when $(t, \delta) \to 0$.
For $v_{n_i}|_{\mathcal{O}_i} \in L^2(\mathcal{O}_i)$, $\|\varsigma(t, v_{n_i}|_{\mathcal{O}_i})\|_{L^2(\mathcal{O})} = \|S_H^{\mathcal{O}}(t)\widetilde{v_{n_i}} - S_D^{\Omega_i}(t)\widetilde{v_{n_i}}\|_{L^2(\mathcal{O})}$ can be bounded above by means of both equations (3) and (4), for $t > 0$ and for all $i \in \{1, 2\}$, by :

$$(4\pi t)^{\frac{-p}{2}} \left\|v_{n_i}|_{\mathcal{O}_i}\right\|^2_{L^2(\mathcal{O}_i)} \left(e^{-\frac{\delta^2}{2t}} Vol(\mathcal{O}_i) + e^{-\frac{\gamma^2}{4t}} Vol(\mathcal{O})\right),$$

with $\gamma = \inf_{x \in \mathcal{O}_j, y \in \mathcal{O}_i} \|x - y\| \geq 2\delta$. ■

Proposition 1 means that the operator $S_H^{\mathcal{O}}(t)$ has 'almost eigenfunctions' which are exactly those of the Dirichlet problem $(\mathcal{P}_\Omega^D)$ restricted to $\mathcal{O}_i \subset \Omega_i$. Moreover, it precises this closeness behaviour with its dependance on t. At last, these eigenfunctions are supported into a single connected component which means that they can be used to characterize if a point belongs to a given Ω_i or not. Finally, if we consider the behaviour of the function $s \mapsto (4\pi s)^{-\frac{p}{2}} e^{-\frac{\delta(y)^2}{4s}}$ involved in (4), we may assume that parameter t should be smaller than $\delta(y)^2$ for all $y \in \mathcal{O}$, and should tend to 0 within a fraction of the distance δ^2 to preserve the asymptotical result of propostion 1.

3 Interpretation of the Spectral Clustering algorithm with Finite Elements

Introduce Lagrange finite elements so that nodes are the set of NJW algorithm data points Σ included in $\bar{\mathcal{O}} \subset \mathbb{R}^p$. Let τ_h be a triangulation on $\bar{\mathcal{O}}$ such that : $h = \max_{K \in \tau_h} h_K$, h_K being a characteristic length of triangle K. Meshing of each part of $\mathcal{O}$ is represented on figure 1 (b). Thus, consider a finite decomposition of the domain : $\bar{\mathcal{O}} = \cup_{K \in \tau_h} K$ in which (K, P_K, Σ_K) satisfies Lagrange finite element assumptions for all $K \in \tau_h$: K is a compact, non empty set of τ_h such that $supp(K) \subset \mathcal{O}$, P_K is a finite dimension vector space generated by shape functions $(\phi_i)_i$ such that $P_K \subset H^1(K)$ and $\Sigma = \cup_{K \in \tau_h} \Sigma_K$.

Let define the finite dimension approximation space: $V_h = \{w \in \mathcal{C}^0(\bar{\mathcal{O}}); \forall K \in \tau_h, w_{|K} \in P_K\}$.
For all $i,\ 1 \leq i \leq n$, let the sequence of shape functions of V_h noted $(\phi_i)_i : \mathcal{O} \to \mathbb{R}$ where $\phi_i = \phi_{|K} \in P_K$ satisfied to: $\phi_i(x_j) = \delta_{ij},\ \forall\ x_j \in \Sigma$, δ_{ij} being the Dirac function of (i, j).
This finite space V_h is generated by the sequence $(\phi_i)_i$ with $V_h = Vect < (\phi_i)_i >\subset L^2(\mathcal{O}, \mathbb{R})$. Let Π_h the linear interpolation from $L^2(\mathcal{O}, \mathbb{R})$ in V_h with the usual scalar product $(\cdot|\cdot)_{L^2}$. According to this notations, the Π_h-mapping of kernel $y \mapsto K_H(t, x - y)$ is defined for

$$\Pi_h(K_H(t, x - y)) = \frac{1}{(4\pi t)^{\frac{p}{2}}} \sum_{i=1}^{n} \exp\left(-\frac{\|x - x_i\|^2}{t}\right) \phi_i(y).$$

Thus, for $t > 0$, the Π_h-mapped operator $S_H^{\mathcal{O}}$ applied to each ϕ_j is, $\forall\ 1 \leq j \leq n$:

$$\Pi_h(S_H^{\mathcal{O}}(t)\phi_j)(x) = \frac{1}{(4\pi t)^{\frac{p}{2}}} \sum_{k=1}^{n} \left(\widehat{A}M\right)_{kj} \phi_k(x) + O(h, x), \tag{7}$$

where M stands for the mass matrix defined by : $M_{ij} = (\phi_i|\phi_j)_{L^2}$ and $\widehat{A} = A + I_n$ where A is the affinity matrix. So, with the same assumptions as in section 2.2, equation (1) can be formulated via the linear interpolation Π_h:

Proposition 2 *Let v_{n_i} be an eigenfunction of $S_D^{\Omega_i}(t)$ in $H^2(\Omega) \cap H_0^1(\Omega)$ associated to the eigenvalue $\exp^{-\lambda_{n_i}t}$ such that $S_D^{\Omega_i}(t)v_{n_i} = \exp^{-\lambda_{n_i}t} v_{n_i}$. Let $i \in \{1, 2\}$, $\widetilde{v_{n_i}} = \mathcal{T}_i(v_{n_i}|_{\mathcal{O}_i})$. Then, for $W_{n_i} = \Pi_h\widetilde{v_{n_i}} \in V_h$ and $t > 0$, :*

$$\Pi_h S_H^{\mathcal{O}}(t)W_{n_i} = \exp^{-\lambda_{n_i}t} W_{n_i} + \psi(t, h), \tag{8}$$

where $\|\psi(t, h)\|_{L^2(V_h)} \to 0$ when $(h, t) \to 0$ and $\delta \to 0$.

Proof sketch. Apply Π_h interpolation to the equation of proposition 1. Using linearity and continuity of Π_h and Cauchy-Schwarz inequality, an upper bound of $\|\Pi_h S_H^{\mathcal{O}}(t)(\Pi_h\widetilde{v_{n_i}} - \widetilde{v_{n_i}})\|_2$ could be given, for a given constant $C_{\Pi_h} > 0$, by :

$$C_{\Pi_h}(4\pi t)^{-\frac{p}{4}} \left\|v_{n_i}|_{\mathcal{O}_i}\right\|_{L^2(\mathcal{O}_i)} Vol(\mathcal{O})^{\frac{3}{2}} \left(e^{-\frac{\delta^2}{2t}} + e^{-\frac{\gamma^2}{4t}}\right)^{\frac{1}{2}}.$$

■

In this approach, we have to estimate the neighbourhood distance δ between Ω_i and $\mathcal{O}_i$ for $i \in \{1, 2\}$. But this result totally relies on both the discretization step h and the Gaussian parameter t. So we investigate their role in the following numerical experiments.

4 Numerical experiments

Let consider two test cases : one in which clusters can be separated by hyperplanes (see fig 2(a)) and another one in which clusters are embedded (see fig 3(a)). We analyze the influence of parameter t on the difference between discretized eigenfunctions of $S_D^{\mathcal{O}}(t)$ and eigenvectors of matrix $\hat{A}M$. In figure 2 (b)-(c)-(d) and figure 3 (b)-(c)-(d), we plot the discretized eigenfunction associated to the first eigenvalue of each connected component. We also indicate in figure 2 (e)-(f)-(g) the product of these discretized eigenfunctions W_{n_i}, for $i \in \{1, 2, 3\}$, with matrix $\hat{A}M$ or with matrix A as well. This shows the locality of eigenfunctions is preserved by the discretized

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

operator $\hat{A}M$ or A. Still, to get better insight about this geometrical invariance property, we finally compare the projection coefficients between these discretized eigenfunctions W_{n_i} with the eigenvectors of $\hat{A}M$ and A that maximizes these projection coefficient. In other words, let $W_{n_i} = \Pi_h \widetilde{v_{n_i}} \in V_h$ and X_l the eigenvector of $\hat{A}M$ such that $l = argmax_j(W_{n_i}|X_j)$, the projection coefficient correponds then to $\omega = \frac{|(W_{n_i}|X_l)|}{\|W_{n_i}\|_2 \|X_l\|_2}$. Since matrix $\hat{A}M$ depends on the parameter t, we indicate in figure (k) how this coefficient ω evolves with varying values of t. In figure (l), we also indicate the difference of the norms $\alpha = \|X_l - W_{n_i}\|_2$. We do the same in figure (m) and (n) with respect to eigenvectors of the affinity matrix A instead of $\hat{A}M$ and noted respectively τ and β. The projection coefficient ω, which gets close to 1 for some optimal value of t, highlights the fact discretized eigenfunctions of Laplacian operator with Dirichlet boundary conditions on some wider subset Ω_i that includes the connected component $\mathcal{O}_i$ are indeed close to eigenvectors of the discretized heat operator without boundary conditions but with a parameter t that needs to be chosen in a given appropriate window.

The vertical dash-dot line in figure (k)-(l)-(m)-(n) indicates the value of the parameter t based on the geometrical point of view proposed in [5]: $t_h = \frac{D_{\max}}{2n^{\frac{1}{p}}}$ with $D_{\max} = \max_{1 \leq i,j \leq n} \|x_i - x_j\|_2$. At a first glance, it is close to an optimum in all these figures. This critical value t_h, seems to give a good estimate of the length δ between clusters. This critical value t_h relies on the assumption that the p-dimensional data set is isotropic enough in the sense that no directions are privileged with very different magnitudes in the distances between points along these directions. It shows the link between eigenvectors and eigenfunctions with respect to parameter t, and that t_h is a good candidate to perform Spectral Clustering.

5 Conclusion

In this paper, we analyze some steps of Spectral Clustering with analogy to eigenvalue problems. From this interpretation, a clustering property based on the eigenvectors and an asymptotic condition on the Gaussian parameter have been extracted. This leads to some interpretation on how spectral clustering works and how results could be affected with respect to the affinity parameter and separation between clusters. However we didn't take into account the normalization step of the spectral clustering algorithm which role is crucial to order the eigenvalues in an appropriate way so that connected components are easy to detect by means of eigenvectors associated to the largest eigenvalues.

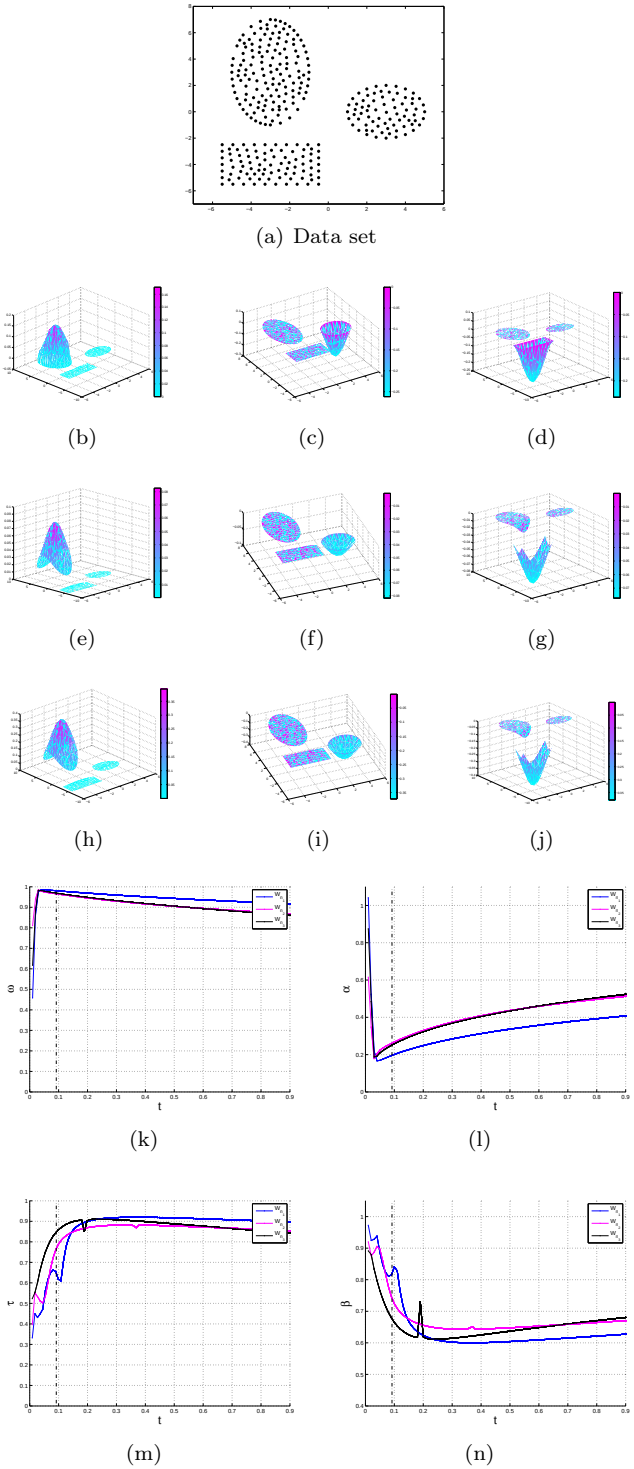


Figure 2: Example 1: (a) Data set with $n = 302$ points, (b), (c), (d) Discretized eigenfunction W_{n_i}, $i \in \{1, 2, 3\}$, (e)-(f)-(g) Product $\hat{A}MW_{n_i}$, (h)-(i)-(j) Product AW_{n_i}, (k)-(l) Correlation ω and Euclidean norm difference α between W_{n_i}, $i \in \{1, 2, 3\}$ and eigenvectors X_l of $\hat{A}M$ functions of t, (m)-(n) Correlation τ and Euclidean norm difference β between W_{n_i}, $i \in \{1, 2, 3\}$ and eigenvectors Y_l of A functions of t.

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

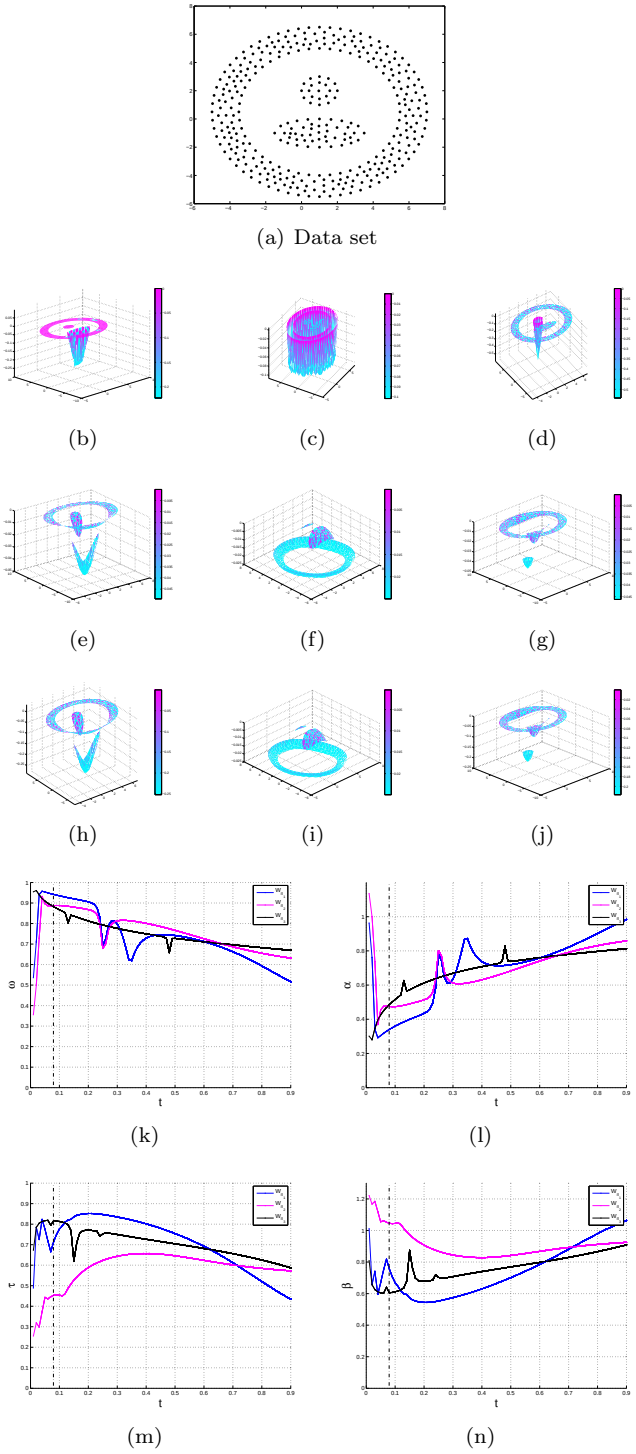


Figure 3: Example 2: (a) Data set with $n = 368$ points, (b), (c), (d) Discretized eigenfunction W_{n_i}, $i \in \{1, 2, 3\}$, (e)-(f)-(g) Product $\hat{A}MW_{n_i}$, (h)-(i)-(j) Product AW_{n_i}, (k)-(l) Correlation ω and Euclidean norm difference α between W_{n_i}, $i \in \{1, 2, 3\}$ and eigenvectors X_l of $\hat{A}M$ functions of t, (m)-(n) Correlation τ and Euclidean norm difference β between W_{n_i}, $i \in \{1, 2, 3\}$ and eigenvectors Y_l of A functions of t.

References

[1] C. Bardos and O. Lafitte. *Une synthese de resultats anciens et recents sur le comportement asymptotique des valeurs propres du Laplacien sur une variete riemannienne.* CMLA, 1998.

[2] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems*, 14(3), 2002.

[3] M. Brand and K. Huang. A unifying theorem for spectral embedding and clustering. *9th International Conference on Artificial Intelligence and Statistics*, 2002.

[4] F. Hirsch and G. Lacombe. *Elements of functional analysis.* Springer, 1999.

[5] S. Mouysset, J. Noailles, and D. Ruiz. Using a Global Parameter for Gaussian Affinity Matrices in Spectral Clustering. In *High Performance Computing for Computational Science: 8th International Conference, Toulouse, France, June 24-27, 2008. Revised Selected Papers*, pages 378–391. Springer, 2009.

[6] B. Nadler and M. Galun. Fundamental limitations of spectral clustering. *Advances in Neural Information Processing Systems*, 19:1017, 2007.

[7] B. Nadler, S. Lafon, R.R. Coifman, and I.G. Kevrekidis. Diffusion maps, spectral clustering and eigenfunctions of fokker-planck operators. *Arxiv preprint math/0506090*, 2005.

[8] A. Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: analysis and an algorithm. *Proc.Adv.Neural Info.Processing Systems*, 2002.

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

[10] U. Von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *Annals of Statistics*, 36(2):555, 2008.

[11] U. Von Luxburg, O. Bousquet, and M. Belkin. Limits of spectral clustering. *Advances in neural information processing systems*, 17, 2004.

[12] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. *Advances in Neural Information Processing Systems*, 17(1601-1608):16, 2004.

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)