Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Spectral Clustering: Interpretation and Gaussian Parameter

Sandrine Mouysset, Joseph Noailles, Daniel Ruiz, and Clovis Tauber

Abstract Spectral clustering consists in creating, from the spectral elements of a Gaussian affinity matrix, a low-dimensional space in which data are grouped into clusters. However, questions about the separability of clusters in the projection space and the choice of the Gaussian parameter remain open. By drawing back to some continuous formulation, we propose an interpretation of spectral clustering with Partial Differential Equations tools which provides clustering properties and defines bounds for the affinity parameter.

1   Introduction

Spectral clustering aims at selecting dominant eigenvectors of a parametrized Gaussian affinity matrix in order to build an embedding space in which the clustering is made. Many interpretations of this method were lead to explain why the clustering is made in the embedding space with graph theory with random walks (Meila and Shi 2001), matrix perturbation theory (Ng et al. 2002), Operators in Manifolds (Belkin and Niyogi 2003), physical models as inhomogeneous ferromagnetic Potts model (Blatt et al. 1996) or Diffusion Maps (Nadler et al. 2006). But all these analysis are investigated asymptotically for a large number of points and do not

S. Mouysset (✉)
University of Toulouse, IRIT-UPS, 118 route de Narbonne, 31062 Toulouse, France
e-mail: sandrine.mouysset@irit.fr

J. Noailles · D. Ruiz
University of Toulouse, IRIT-ENSEEIHT, 2 rue Camichel, 31071 Toulouse, France
e-mail: joseph.noailles@irit.fr; daniel.ruiz@irit.fr

C. Tauber
University of Tours, Hopital Bretonneau, 2 boulevard Tonnelle, 37044 Tours, France
e-mail: clovis.tauber@univ-tours.fr

explain why this method works for a finite data set. Moreover, another problem still arise: the affinity parameter influences the clustering results (Ng et al. 2002; Von Luxburg 2007). And the difficulty to define an adequate parameter seems to be slightly connected to the lack of some clustering property explaining how the grouping in this low-dimensional space correctly defines the partitioning in the original data.

In this paper, we propose a fully theoretical interpretation of spectral clustering whose first steps were introduced by Mouysset et al. (2010). From this, we define a new clustering property in the embedding space at each step of the study and new results showing the rule of the Gaussian affinity parameter. After recalling the spectral clustering method and the rule of the affinity parameter in Sect. 2.1, we propose a continuous version of the Spectral Clustering with Partial Differential Equations (PDE). To do so, we consider a sampling of connected components and, from this, we draw back to original shapes. This leads to formulate spectral clustering as an eigenvalue problem where data points correspond to nodes of some finite elements discretization and to consider the Gaussian affinity matrix A as a representation of heat kernel and the affinity parameter σ as the heat parameter t. Hence, the first step is to introduce an eigenvalue problem based on heat equation which is defined with a Dirichlet boundary problem. From this, in Sect. 2.2, we deduce an "almost" eigenvalue problem which can be associated to the Gaussian values. Thus identifying connected component appears to be linked to these eigenfunctions. Then, by introducing the Finite Elements approximation and mass lumping, we prove in Sect. 2.3 that this property is preserved with conditions on t when looking at eigenvectors given by spectral clustering algorithm. Finally, in Sect. 3, we study numerically the difference between eigenvectors from the spectral clustering algorithm and their associated discretized eigenfunctions from heat equation on a geometrical example, as a function of the affinity parameter t.

2 Interpretation

In the following, spectral clustering and its inherent problem are presented. Then we propose a continuous version of this method.

2.1 *Spectral Clustering: Rule of Gaussian Parameter*

Let consider a data set $\mathcal{P} = \{x_i\}_{i=1..N} \in \mathbb{R}^p$. Assume that the number of targeted clusters k is known. First, the spectral clustering consists in constructing the parametrized affinity matrix based on the Gaussian affinity measure between points of the data set $\mathcal{P}$. After a normalization step, by stacking the k largest eigenvectors, the spectral embedding in $\mathbb{R}^k$ is created. Each row of this matrix represents a data point x_i which is plotted in this embedding space and then grouped into clusters via
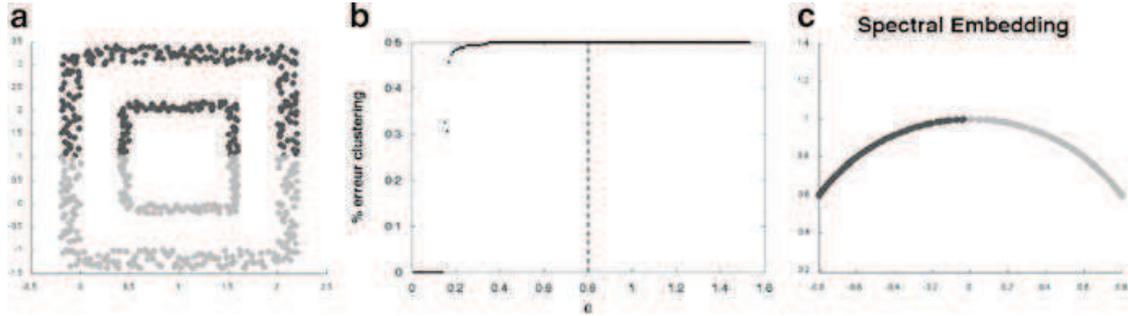
Fig. 1 Geometrical example: (**a**) clustering result for $\sigma = 0.8$, (**b**) percentage of clustering error function of σ, (**c**) spectral embedding space for $\sigma = 0.8$

the K-means method. Finally, thanks to an equivalence relation, the final partition of data set is directly defined from the clustering in the embedding space.

So this unsupervised method is mainly based on the Gaussian affinity measure, its parameter σ and its spectral elements. Moreover, it is known that the Gaussian parameter conditions the separability between the clusters in the spectral embedding space and should be well chosen (Von Luxburg 2007). The difficulty to fix this choice seems to be tightly connected to the lack of results explaining how the grouping in this low-dimensional space defines correctly the partitioning in the original data for a finite data set. Figure 1 summaries these previous remarks via a percentage of clustering which evaluated the percentage of mis-clustered points applied on a geometrical example of two concentric rectangles (Fig. 1a). For $\sigma = 0.8$, value which provides clustering errors (Fig. 1b), the two clusters defined with K-means are represented in the spectral embedding (Fig. 1c) by the respective black and grey colors. A piece of circle in which no separation by hyperplane is possible is described. Thus, in the original space, both rectangles are cut in two and define a bad clustering as show in Fig. 1a.

2.2 Through an Interpretation with PDE Tools

As spectral elements used in spectral clustering 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 as shown in Fig. 2. In that way, we first have to define a clustering compatibility which establishes the link between continuous interpretation and the discrete case. So we consider an open set Ω subdivided by k disjoints connected components of Ω.

Definition 1 (Clustering Compatibility). Let Ω be a bounded open set in $\mathbb{R}^p$ made by Ω_i, $i \in 1, .., k$ disjoint connected components such that: $\Omega = \bigcup_{i=1}^{k} \Omega_i$. Let $\mathcal{P}$ be a set of points $\{x_i\}_{i=1}^{N}$ in the open set Ω. Let note $\mathcal{P}_j$, for $j = \{1, .., k\}$, the non empty set of points of $\mathcal{P}$ in the connected component Ω_j of Ω: $\mathcal{P}_j = \Omega_j \cap \mathcal{P}, \forall j \in \{1, .., k\}$. Let $C = \{C_1, .., C_{k'}\}$ be a partition of $\mathcal{P}$.
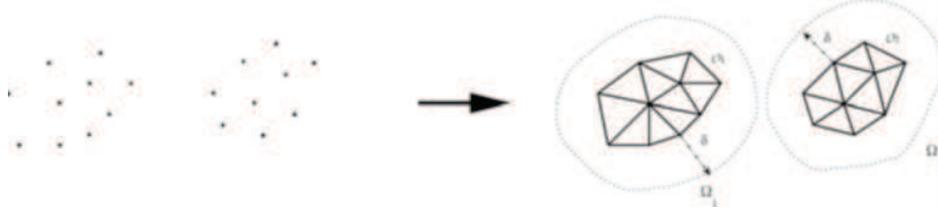
Fig. 2 Principle of the interpretation with PDE tools

Suppose that $k = k'$ then C is a compatible clustering if $\forall j = \{1, .., k'\}, \exists i \in \{1, .., k\}, C_j = \mathcal{P}_i$.

To make a parallel version in the $L^2(\Omega)$ space, data points which believe in a subset of Ω are equivalent to believe in the same connected component. In the following, we will formulate spectral clustering as an eigenvalue problem by assuming data points as nodes of some finite elements discretization and by considering Gaussian affinity matrix as a representation of heat kernel. But as the spectrum of heat operator in free space is essential, we will make a link with a problem defined on bounded domain in which the spectrum is finite. Then, due to the fact that we compare the discrete data defined by the elements of the affinity matrix with some L^2 functions which are the solutions of heat equation, we will introduce an explicit discretization with the Finite Element theory and the mass lumping to cancel all knowledge about the mesh. Then we will make a feedback of this analysis for the application of spectral clustering by defining clustering properties following the successive approximations. Finally, this study will lead to a functional rule of σ and a new formulation of a spectral clustering criterion.

2.2.1 Link Between Gaussian Affinity and Heat Kernel in $\mathbb{R}^p$

Let recall the Gaussian affinity element A_{ij} between two data points x_i and x_j is defined by $A_{ij} = \exp\left(-\|x_i - x_j\|^2/2\sigma^2\right)$. A direct link between the affinity A_{ij} and the heat kernel on $\mathbb{R}_+^* \times \mathbb{R}^p$, defined by $K_H(t, x) = (4\pi t)^{-\frac{p}{2}} \exp\left(-\|x\|^2/4t\right)$ could be established as follows:

$$A_{ij} = (2\pi\sigma^2)^{\frac{p}{2}} K_H\left(\sigma^2/2, x_i - x_j\right), \ \forall i \neq j, \ \forall (i, j) \in \{1, .., N\}. \tag{1}$$

Equation (1) permits defining the affinity measure as a limit operator: the Gaussian affinity is interpreted as the heat kernel of a parabolic problem and its Gaussian parameter σ as a heat parameter t. Consider the following parabolic problem which is called heat equation, for $f \in L^2(\mathbb{R}^p)$:

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

Due to the fact that the spectrum of heat operator in free space, noted S_H, is essential and eigenfunctions are not localized in $\mathbb{R}^p$ without boundary conditions, we have to restrict the domain definition and make a link with a problem on a bounded domain Ω in which the eigenfunctions could be studied.

2.2.2 Clustering Property with Heat Equation

Let now introduce the initial value problem in $L^2(\Omega)$, for $f \in L^2(\Omega)$:

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

Denote by K_D the Green's kernel of $(\mathcal{P}_\Omega)$. The solution operator in $H^2(\Omega) \cap H_0^1(\Omega)$ associated to this problem is defined, for $f \in L^2(\Omega)$, by:

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

Let consider $\{(\widetilde{v_{n,i}})_{n,i>0}, \ i \in \{1, .., k\}\} \in H_0^1(\Omega)$ such that $(\widetilde{v_{n,i}})_{n,i>0}$ are the solutions of $\Delta\widetilde{v_{n,i}} = \lambda_{n,i}\widetilde{v_{n,i}}$ on Ω_i for $i \in \{1, .., k\}$ and $n > 0$ and extend $\widetilde{v_{n,i}} = 0$ on $\Omega\backslash\Omega_i$. These functions are eigenfunctions of $(\mathcal{P}_\Omega)$ and the union of these eigenfunctions is an Hilbert basis of $H_0^1(\Omega)$. Moreover, as $\Omega = \bigcup_{i=1}^k \Omega_i$, for all $i \in \{1, .., k\}$ and $n > 0$, the eigenfunctions, noted $\{(\widetilde{v_{n,i}})_{n,i>0}, \ i \in \{1, .., k\}\}$, satisfied: $S_D(t)\widetilde{v_{n,i}} = e^{-\lambda_{n,i}t}\widetilde{v_{n,i}}$. So the eigenfunctions of S_D have a geometrical property: its support is included in only one connected component. Thus a clustering property in the spectral embedding space could be established.

Proposition 1 (Clustering Property). *For all point $x \in \Omega$ and $\epsilon > 0$, let note ρ_x^ϵ a regularized Dirac function centred in x: $\rho_x^\epsilon \in C^\infty(\Omega, [0,1])$, $\rho_x^\varepsilon(x) = 1$ and $supp(\rho_x^\varepsilon) \subset \mathcal{B}(x, \varepsilon)$. The eigenfunctions of S_D, noted $\widetilde{v_{n,i}}$, for $i \in \{1, .., k\}$ and $n > 0$ such that for all $x \in \Omega$ and all $i \in \{1, .., k\}$ and for all $t > 0$, the following result is satisfied:*

$$\left[\exists \varepsilon_0 > 0, \ \forall \varepsilon \in]0, \varepsilon_0[, \ \exists n > 0, \ (S_D(t)\rho_x^\varepsilon | \widetilde{v_{n,i}})_{L^2(\Omega)} \neq 0\right] \Longleftrightarrow x \in \Omega_i \qquad (2)$$

where $(f|g)_{L^2(\Omega)} = \int_\Omega f(y)g(y)dy, \forall(f, g) \in L^2(\Omega)$ is the usual scalar product in L^2.

Proof. By contrapositive, let $i \in \{1, .., k\}$ and a point $x \in \Omega_j$ with any $j \neq i$. Let $d_x = d(x, \partial\Omega_j) > 0$ be the distance of x from the boundary of Ω_j. According to the hypothesis on Ω, we have $d_0 = d(\Omega_i, \Omega_j) > 0$. So for all $\varepsilon \in]0, \inf(d_x, d_0)[$, $\mathcal{B}(x, \varepsilon) \subset \Omega_j$. Then for all $t > 0$, $supp(S_D(t)\rho_x^\varepsilon) \subset \Omega_j$ and so, for $n > 0$,

$(S_D(t)\rho_x^\varepsilon|\widetilde{v_{n,i}})_{L^2(\Omega)} = 0$. So there does not any $\varepsilon_0 > 0$ which verifies the direct implication of (2). Reversely, let $x \in \Omega_i$ and $\varepsilon \in]0, \inf(d_x, d_0)[$, $\mathcal{B}(x, \varepsilon) \subset \Omega_i$. So the support of ρ_x^ε is in Ω_i. As the $(\widetilde{v_{n,i}})_{n>0}$ is an Hilbert basis of $L^2(\Omega_i)$ and that $\rho_x^\varepsilon(x) = 1 \neq 0$ then there exists $n > 0$ such that $(\rho_x^\varepsilon|\widetilde{v_{n,i}}) \neq 0$. In this case, $(S_D(t)\rho_x^\varepsilon|\widetilde{v_{n,i}})_{L^2(\Omega)} = e^{-\lambda_{n,i}t}(\rho_x^\varepsilon|\widetilde{v_{n,i}}) \neq 0$.

By considering an open subset O which approximates from the interior the open set Ω such that $Volume(\Omega \backslash O) \leq \epsilon$, for $\epsilon > 0$, both heat operators of $(\mathcal{P}_{\mathbb{R}^p})$ and $(\mathcal{P}_\Omega)$ could be compared in O. Let δ be the distance from O to Ω as shown in Fig. 2. Due to the fact that the difference between the Green kernels K_H and K_D could be estimated in O and is function of the heat parameter t, the geometrical property could thus be preserved on the heat operator in free space restricted to O. Let $v_{n,i}$ be the eigenfunction $\widetilde{v_{n,i}}$ which support is restricted to O, for all $i \in \{1, .., k\}$ and $n > 0$. From this, we obtain, for $0 < t < \delta^2$:

$$S_H^O(t)v_{n,i} = \exp^{-\lambda_{n_i}t} v_{n,i} + \eta(t, v_{n,i}), \tag{3}$$

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

So we can prove that on O, the eigenfunctions for the solution operator for bounded heat equation are quasi-eigenfunctions for S_H^O plus a residual (Mouysset et al. 2010). The clustering property adapted to the restricted heat operator S_H^O remains introducing an hypothesis on the heat parameter t. Moreover, (2) is modified with non-null values by introducing a gap between scalar product with eigenfunctions such that for all $x > 0$ and all $i \in \{1, .., k\}$:

$$\begin{bmatrix} \exists \varepsilon_0 > 0, \ \exists \alpha > 0, \ \forall \varepsilon \in]0, \varepsilon_0[, \ \exists n > 0, \ \forall t > 0 \text{ small enough,} \\ v_{n,i} = \arg\max_{\{v_{m,j}, m \in \mathbb{N}, j \in [|1,k|]\}} \left| (S_H^O(t)\rho_x^\varepsilon|v_{m,j})_{L^2(O)} \right| \\ \text{and } \left| (S_H^O(t)\rho_x^\varepsilon|v_{n,i})_{L^2(O)} \right| > \alpha \end{bmatrix} \iff x \in O_i. \tag{4}$$

These previous results prove that in infinite dimension, a clustering could be realized in the spectral embedding space because the eigenfunctions have a geometrical property. This study leads to the following question: do eigenvectors of the affinity matrix behave like eigenfunctions of $(\mathcal{P}_\Omega)$?

2.3 Discretization with Finite Elements

From this, we will look for a similar behaviour onto eigenvectors of A by introducing a finite dimension representation matching with the initial data set $\mathcal{P}$ with help of the finite elements (Ciarlet 1978). So, we consider data points as finite dimensional approximation and elements of the affinity matrix built from data points as nodal values of S_H^O.

2.3.1 Approximation in Finite Dimension

Let τ_h be a triangulation on $\bar{O}$ such that: $h = \max\limits_{K \in \tau_h} h_K$, h_K being a characteristic length of triangle K. Let consider a finite decomposition of the domain: $\bar{O} = \cup_{K \in \tau_h} K$ in which (K, P_K, Σ_K) satisfies Lagrange finite element assumptions for all $K \in \tau_h$. We define also the finite dimension approximation space: $V_h = \{w \in C^0(\bar{O}); \forall K \in \tau_h, w_{|K} \in P_K\}$ and denote Π_h the linear interpolation from $C^0(\bar{O})$ in V_h with the usual scalar product $(\cdot|\cdot)_{L^2(V_h)}$ (Ciarlet 1978). According to this notations, for $t > 0$, the Π_h-mapped operator S_H^O applied to each shape function ϕ_j is, for $h^{3p+2} < t^2$, for all $1 \le j \le N$:

$$(4\pi t)^{\frac{p}{2}} \Pi_h(S_H^O(t)\phi_j)(x) = \sum_{k=1}^{N} ((A + \mathbb{I}_N)M)_{kj} \, \phi_k(x) + O\left(\frac{h^{3p+2}}{t^2}\right), \quad (5)$$

where M stands for the mass matrix defined by: $M_{ij} = (\phi_i|\phi_j)_{L^2(V_h)}$. Equation (5) means that the affinity matrix defined in (1) in spectral algorithm is interpreted as the Π_h-projection of operator solution of $(\mathcal{P}_{\mathbb{R}^p})$ with M mass matrix from Finite Element theory (Mouysset et al. 2010).

So we could formulate finite elements approximation of continuous clustering result (3). From the eigenfunctions of S_D restricted to O, their projection in V_h, noted $W_{n,i}$, are defined by: $W_{n,i} = \Pi_h v_{n,i} \in V_h, \forall i \in \{1, .., k\}$. So, for $h^{\frac{3p+2}{2}} < t < \delta^2$, the following result could be established:

$$(4\pi t)^{\frac{-p}{2}} (A + \mathbb{I}_N)M W_{n,i} = e^{-\lambda_{n,i}t} W_{n,i} + \Psi(t, h), \quad (6)$$

where $\|\Psi(t,h)\|_{L^2(V_h)} \to 0$ and $\delta \to 0$. Equation (6) shows that the geometrical property is preserved in finite dimension on the eigenvectors of $(A + \mathbb{I}_N)M$. Moreover, a lower bound for the heat parameter was defined. But all this previous results include the mass matrix which is totally dependent of the finite elements. In order to cancel this dependence, mass lumping process is investigated.

2.3.2 Mass Lumping

The mass lumping method consists in using a quadrature formula whose integration points are the interpolation points of the finite element. So let $\mathcal{I}_k$ be the list of indices of points which are element of $K \in \tau_h$. Let consider the quadrature scheme exact for polynomials of degree ≤ 1:

$$\int_K \phi(x)dx \approx \sum_{k \in \mathcal{I}_k} \frac{|K|}{3}\phi(x_{i_k}) \quad (7)$$

where $|K|$ is the area of the finite element K. So, with additional regularity condition on the mesh which bounds $|K|$, the mass lumping permits considering the mass matrix M as a homogeneous identity matrix. So (6) is modified so that, $\exists\, \alpha > 0$, such that:

$$\alpha\,(A + \mathbb{I}_N)\,W_{n,i} = e^{-\lambda_{n,i}t}\,W_{n,i} + \Psi'(t, h), \tag{8}$$

where $\|\Psi'(t, h)\|_{L^2(V_h)} \to 0$ and $\delta \to 0$. The approximation in finite dimension of the clustering property (4) is reformulated as follows, for all $x_r \in \mathcal{P}$, for all $i \in \{1, .., k\}$:

$$\left[\begin{array}{l} \exists \alpha > 0,\ \exists n > 0,\ \forall t > 0, t, h^2/t \text{ and } h^{(3p+1)}/t^2 \text{ small enough,} \\ W_{n,i} = \arg\max_{\{W_{m,j}, m \in \mathbb{N}, j \in [|1,k|]\}} \left|\left((A + \mathbb{I}_N)_{.r}\,|W_{m,j}\right)_{L^2(V_h)}\right| \\ \text{and } \left|\left((A + \mathbb{I}_N)_{.r}\,|W_{n,i}\right)_{L^2(V_h)}\right| > \alpha \end{array}\right] \iff x_r \in O_i,$$

$$\tag{9}$$

where $((A + \mathbb{I}_N))_{.r}$ is the rth column of the matrix $(A + \mathbb{I}_N)$, for all $r \in \{1, ..N\}$. This leads to the same clustering for a set of data points either we consider eigenfunctions in $L^2(\Omega)$ or Π_h-interpolated eigenfunction in the approximation space V_h. With an asymptotic condition on the heat parameter t (or Gaussian parameter σ), points which are elements of the same cluster have the maximum of their projection coefficient along the same eigenvector. So the clustering in spectral embedding space provides the clustering in data space.

3 Gaussian Parameter: A Geometrical Example

This previous theoretical interpretation proves that the Gaussian parameter should be chosen within a specific interval in order to improve the separability between clusters in the spectral embedding space. In order to experiment the parallel between continuous version and the approximate one, we consider a geometrical example with non convex shapes as shown in Fig. 3a. For each connected component (or each cluster) $i \in \{1, 2\}$, the discretized eigenfunction, noted $W_{1,i}$, associated to the first eigenvalue of each connected component and the eigenvectors, noted Y_i, which gives maximum projection coefficient with $W_{1,i}$ are respectively plotted in Fig. 3b, c and e, f. The correlation ω between $W_{1,i}$ and Y_i is represented as a function of the heat parameter t in Fig. 3d: $\omega = |(W_{1,i}|Y_i)|(\|W_{1,i}\|_2\|Y_i\|_2)^{-1}$. The vertical black dash dot lines indicate the lower and upper estimated bounds of the heat parameter. In this interval, the correlation between the continuous version and the eigenvectors of the Gaussian affinity matrix is maximum. So the clusters are well separated in the spectral embedding space.
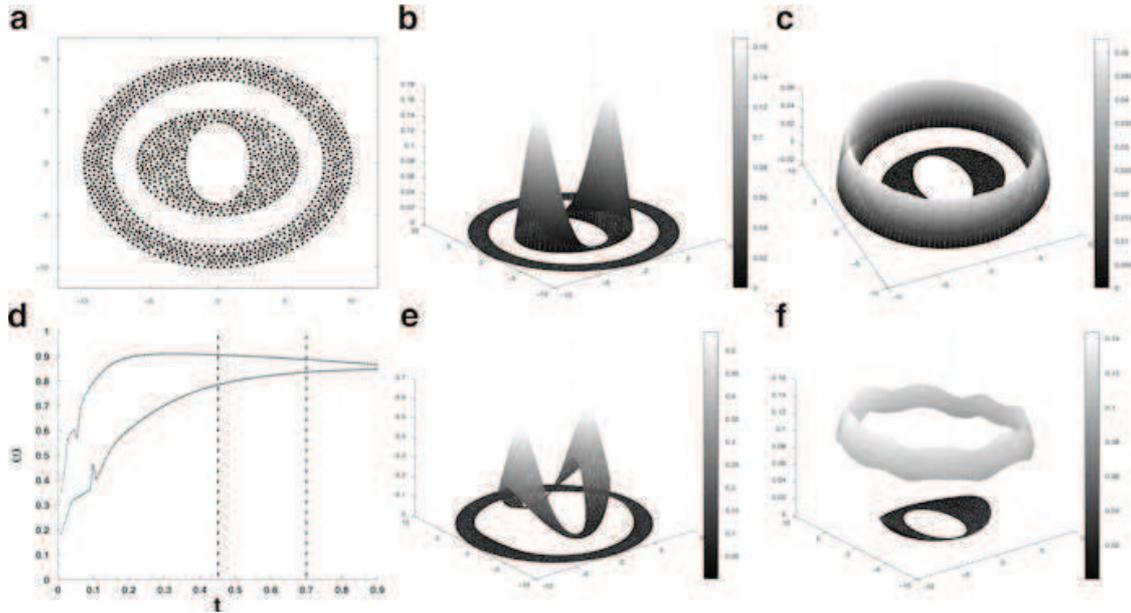
Fig. 3 (**a**) Data set ($N = 669$), (**b**) and (**c**) discretized eigenfunctions of S_D, (**d**) correlation ω between the continuous version and its discrete approximation function of t, (**e**) and (**f**) eigenvectors from A which provides the maximum projection coefficient with the eigenfunctions of S_D

4   Conclusion

In this paper, spectral clustering was formulated as an eigenvalue problem. From this interpretation, a clustering property on the eigenvectors and some conditions on the Gaussian parameter have been defined. This leads to understand how spectral clustering works and to show how clustering results could be affected with a bad choice of the affinity parameter. But we do not take into account the normalization step in the whole paper but its rule is crucial for ordering largest eigenvectors for each connected components to the first eigenvectors and should be studied.

References

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation, 15*(6), 1373–1396.

Blatt, M., Wiseman, S., & Domany, E. (1996). Superparamagnetic clustering of data. *Physical Review Letters, 76*(18), 3251–3254.

Ciarlet, P. G. (1978). *The finite element method for elliptic problems. Series studies in mathematics and its applications* (Vol. 4). Amsterdam: North-Holland.

Meila, M., & Shi, J. (2001). A random walks view of spectral segmentation. In *Proceedings of eighth international workshop on artificial intelligence and statistics (AISTATS) 2001.*

Mouysset, S., Noailles, J., & Ruiz, D. (2010). On an interpretation of spectral clustering via heat equation and finite elements theory. In *Proceedings of international conference on data mining and knowledge engineering (ICDMKE)* (pp. 267–272). Newswood Limited.

Nadler, B., Lafon, S., Coifman, R. R., & Kevrekidis, I. G. (2006). Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis: Special Issue on Diffusion Maps and Wavelets, 21*(1), 113–127.

Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). *On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems* (pp. 849–856). Cambridge: MIT.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing, 17*(4), 395–416. Berlin: Springer.