Séminaire de Calcul Scientifique du CERMICS

Geometric Structure of Graph Laplacian Embeddings

Nicolás García Trillos (Brown University)

31 mai 2018

Geometric Structure of Graph Laplacian Embeddings.

Nicolás García Trillos
Brown University

CERMICS
Ecole des Ponts ParisTech
May 31st 2018

Outline

1. Spectral clustering.
2. Geometry of spectral embeddings.
3. Discrete to continuum limit of graph-based procedures of machine learning.

This presentation mostly based on:

1. *Error estimates for spectral convergence of the graph Laplacian on random geometric graphs towards the Laplace–Beltrami operator* (2018) with M. Gerlach, M. Hein, D. Slepcev.
2. *Geometric structure of graph Laplacian embeddings* (In preparation) with F. Hoffman and B. Hosseini.

What is clustering and what is spectral clustering?

Given a data set $X = \{x_1, \dots, x_n\}$:

find meaningful **clusters** $A_1, \ldots, A_N$ in the data set:

One may have a similarity matrix W to achieve this.

What is spectral clustering?

1. **Input:** Similarity graph (X, W).
2. **Output:** Clusters $A_1, \dots, A_N$

Two steps:

1. Embedding step $\mathcal{F}_n : X \to \mathbb{R}^N$
2. N-means on $\mathcal{F}_n(x_1), \dots, \mathcal{F}_n(x_n)$.

Embedding step

Consider the discrete differential operator:

$$\Delta_n := D - W,$$

i.e. **graph Laplacian**.
Self-adjoint with respect to $\langle \cdot, \cdot \rangle_{\nu_n}$ and with associated Dirichlet energy

$$\sum_{ij} W_{ij} |u_n(x_i) - u_n(x_j)|^2$$

Let $u_{1,n}, \dots, u_{N,n}$ be first N eigenfunctions of Δ_n.

$$\mathcal{F}_n : x_i \in \mathcal{M} \longmapsto \begin{pmatrix} u_{1,n}(x_i) \\ \vdots \\ u_{N,n}(x_i) \end{pmatrix}$$

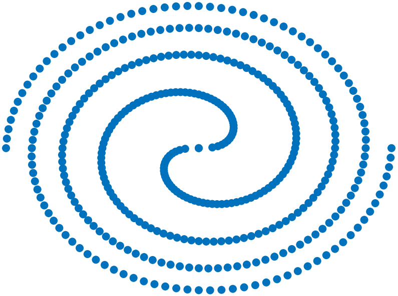
In this talk:

- $x_1, \dots, x_n$ i.i.d. draws from some distribution $d\nu = \rho dx$ on $\mathcal{M} \subseteq \mathbb{R}^d$.
- W_{ij} obtained as follows:

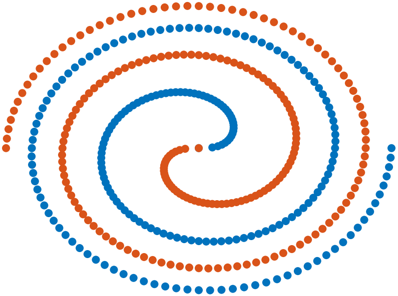
$$\tilde{W}_{ij} := \eta\left(\frac{|x_i - x_j|}{\varepsilon_n}\right)$$

$$W_{ij} := \frac{\tilde{W}_{ij}}{\sqrt{\tilde{d}_i}\sqrt{\tilde{d}_j}}$$

Why not using N-means directly?



Why not using N-means directly?



Motivating trivial example: Suppose $\mathcal{M}$ has N connected components:

$$\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_N$$

For a well chosen value of ε_n, (X, W) will have N connected components $A_{1,n}, \ldots, A_{N,n}$.

Geometry of spectral embeddings

In more generality: let us assume that $x_1, \dots, x_n$ are samples from mixture model:

$$\rho(x) := \sum_{k=1}^{N} w_k \rho_k(x), \quad x \in \mathcal{M}.$$

Goal: Describe the geometry of the spectral embedding $\mathcal{F}_{n\sharp}\nu_n$ when the components of mixture model are **well separated**.

Informal main result

Provided mixture model is well separated, with very high probability, there are numbers (α, δ, r) s.t:

$$\mathcal{F}_{n\sharp}\nu_n \left(\bigcup_{j=1}^{N} C(e_j, \alpha, r) \right) \geq 1 - \delta,$$

where $C(e_j, \alpha, r)$ is the set

$$C(e_j, \alpha, r) := \left\{ z \in \mathbb{R}^N \ : \ \frac{z \cdot e_j}{|z|} > \cos(\alpha), \quad |z| > r \right\}.$$

First, look at the continuum setting.

Consider the differential operator:

$$\Delta_\rho(u) := -\frac{1}{\rho} div(\rho \nabla u)$$

Self-adjoint with respect to $\langle \cdot, \cdot \rangle_\rho$ and with associated Dirichlet energy

$$\int_{\mathcal{M}} |\nabla u|^2 \rho dx$$

Let $u_1, \dots, u_N$ be first N eigenfunctions of Δ_ρ.

$$\mathcal{F} : x \in \mathcal{M} \longmapsto \begin{pmatrix} u_1(x) \\ \vdots \\ u_N(x) \end{pmatrix}$$

We will show that when ρ mixture model is **well separated** then $F_\sharp \nu$ has an **orthogonal cone structure**.

Orthogonal cone structure

Definition

$\mu \in \mathcal{P}(\mathbb{R}^N)$ has orthogonal cone structure with parameters (α, δ, r) if

$$\mu\left(\bigcup_{j=1}^{N} C(e_j, \alpha, r)\right) \geq 1 - \delta,$$

where $C(e_j, \alpha, r)$ is the set

$$C(e_j, \alpha, r) := \left\{ z \in \mathbb{R}^N \ : \ \frac{z \cdot e_j}{|z|} > \cos(\alpha), \quad |z| > r \right\}.$$

Well separated mixture models

This notion is defined in terms of three quantities:

1. $\mathcal{S}$: overlapping parameter.
2. $\mathcal{C}$: coupling parameter.
3. Λ: indivisibility parameter.

Overlapping parameter:

$$\mathcal{S} := \max_{i \neq j} \int_{\mathcal{M}} \frac{\rho_i \rho_j}{\rho} dx, \tag{1}$$

Coupling parameter:

$$\mathcal{C} := \max_k \mathcal{C}_k$$

where

$$\mathcal{C}_k := \frac{1}{4}\int_{\mathcal{M}} \left| \frac{\nabla \rho_k}{\rho_k} - \frac{\nabla \rho}{\rho} \right|^2 \rho_k dx = \frac{1}{4}\text{Fisher}_\rho(\rho_k), \quad k = 1, \dots, N.$$

Indivisibility parameter:

$$\Lambda := \min_{k=1,\dots,N} \Lambda_k, \tag{2}$$

where

$$\Lambda_k := \min_{u \perp \mathbb{1}} \frac{\int_{\mathcal{M}} |\nabla u|^2 \rho_k(x)}{\langle u, u \rangle_{\rho_k}}.$$

Definition well separated mixture model:

Informally,

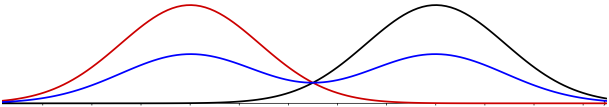
$$\rho(x) := \sum_{k=1}^{N} w_k \rho_k(x), \quad x \in \mathcal{M}.$$

is a well separated mixture model if:

$$\mathcal{C}, \mathcal{S} \ll \Lambda$$

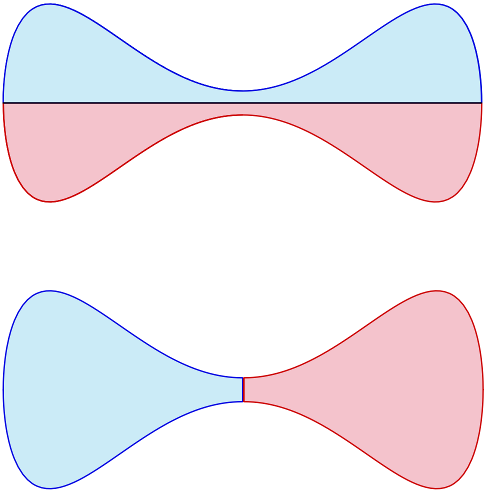
$$\mathcal{S} \ll 1$$

Example 1: Mixture of Gaussians



$$\rho_1(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \rho_2(x) := \frac{1}{\sqrt{2\pi}} e^{-(x-\gamma)^2/2}.$$

Example 2: Dumbbell



Introduce the functions:

$$q_k(x) := \sqrt{\frac{\rho_k(x)}{\rho(x)}}, \quad x \in \mathcal{M}$$

And consider the map F^Q:

$$\mathcal{F}^Q : x \in \mathcal{M} \longmapsto \begin{pmatrix} q_1(x) \\ \vdots \\ q_N(x) \end{pmatrix}$$

- If $\mathcal{S}$ is small then $\mathcal{F}_{\sharp}^{Q}\nu$ has orthogonal cone structure.
- If Λ is much larger than $\mathcal{C}$ and $\mathcal{S}$, and if $\mathcal{S}$ is sufficiently small then $\mathcal{F}_{\sharp}\nu$ also has an orthogonal cone structure.

Analysis for $\mathcal{F}_{n\sharp}\nu_n$

$$\mathcal{F}_n : x_i \longmapsto \begin{pmatrix} u_{1,n}(x_i) \\ \vdots \\ u_{N,n}(x_i) \end{pmatrix}, \quad \mathcal{F} : x \in \mathcal{M} \longmapsto \begin{pmatrix} u_1(x) \\ \vdots \\ u_N(x) \end{pmatrix}$$

To show that

$$d_2(\mathcal{F}_{n\sharp}\nu_n, \mathcal{F}_\sharp \nu) \ll 1$$

would like to show that as $n \to \infty$, $u_{i,n}$ converges to u_i in a convenient sense...

$$\mathcal{F}_n : x_i \longmapsto \begin{pmatrix} u_{1,n}(x_i) \\ \vdots \\ u_{N,n}(x_i) \end{pmatrix}, \quad \mathcal{F} : x \in \mathcal{M} \longmapsto \begin{pmatrix} u_1(x) \\ \vdots \\ u_N(x) \end{pmatrix}$$

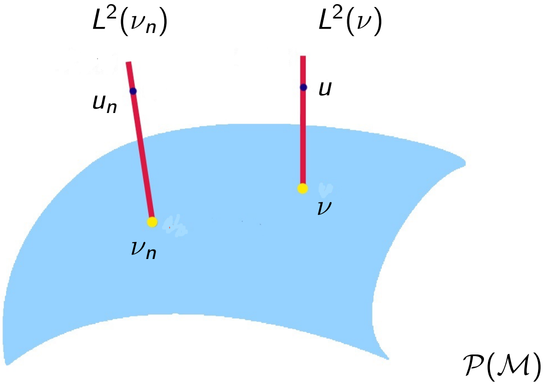
To show that

$$d_2(\mathcal{F}_{n\sharp}\nu_n, \mathcal{F}_{\sharp}\nu) \ll 1$$

would like to show that as $n \to \infty$, $u_{i,n}$ converges to u_i in a convenient sense...

in what sense?

TL^2 space



$$TL^2 = \{(\theta, v) \,:\, \theta \in \mathcal{P}(\mathcal{M}), \quad v \in L^2(\theta)\}.$$

with distance between (θ_1, v_1) and (θ_2, v_2):

$$\inf_{\vartheta \in \Gamma(\theta_1,\theta_2)} \int_{\mathcal{M}\times\mathcal{M}} d^2_{\mathcal{M}}(x,y) d\vartheta(x,y) + \int_{\mathcal{M}\times\mathcal{M}} |v_1(x) - v_2(y)|^2 d\vartheta(x,y).$$

Theorem (NGT, Gerlach , Hein , Slepcev 18')

With very high probability, there exists a transport map $T_n : \mathcal{M} \to X_n$ with $T_{n\sharp}\nu = \nu_n$ such that

$$\sup_{x\in\mathcal{M}} d_{\mathcal{M}}(x, T_n(x)) \sim \frac{(\log(n))^{p_m}}{n^{1/m}}$$

Moreover, for $i = 1, \dots, N$

$$\int_{\mathcal{M}} |u_i(x) - u_{i,n}(T_n(x))|^2 d\nu(x) \lesssim C\left(\frac{\varepsilon_n n^{1/m}}{(\log(n))^{p_m}} + \varepsilon_n\right),$$

Notice that:

$$d_2(\mathcal{F}_\sharp\nu, \mathcal{F}_{n\sharp}\nu_n) \le \int_{\mathcal{M}} |\mathcal{F}(x) - \mathcal{F}_n(T_n(x))|^2 d\nu(x).$$

- If $\mathcal{S}$ is small then $\mathcal{F}^{Q}_{\sharp}\nu$ has orthogonal cone structure.
- If Λ is much larger than $\mathcal{C}$ and $\mathcal{S}$, and if $\mathcal{S}$ is sufficiently small then $\mathcal{F}_{\sharp}\nu$ also has an orthogonal cone structure.
- If there is a theorem providing rates of convergence of $u_{i,n}$ towards u_i, then $\mathcal{F}_{n\sharp}\nu_n$ also has orthogonal cone structure.

Some references studying convergence of discrete to continuum eigenfunctions

- M. Belkin and P. Niyogi, Convergence of Laplacian eigenmaps, Advances in Neural Information Processing Systems (NIPS), 19 (2007), p. 129.
- L. Hagen and A. Kahng, New spectral methods for ratio cut partitioning and clustering., IEEE Trans. Computer-Aided Design, 11 (1992), pp. 1074-1085
- J. Shi and J. Malik, Normalized Cuts and Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 22(8) (2000), pp. 888-905.
- D. Ting, L. Huang, and M. I. Jordan, An analysis of the convergence of graph Laplacians, in Proceedings of the 27th International Conference on Machine Learning, 2010.
- U. von Luxburg, A tutorial on spectral clustering, Statistics and computing, 17 (2007), pp. 395-416.

- U. von Luxburg, M. Belkin, and O. Bousquet, Consistency of spectral clustering, Ann. Statist., 36 (2008), pp. 555-586.
- X. Wang, Spectral convergence rate of graph laplacian, arXiv preprint arXiv:1510.08110, (2015).
- N. Garcia Trillos and D. Slepcev, A variational approach to the consistency of spectral clustering, ACHA (2016).
- **G. Schiebinger, M. J. Wainwright, and B. Yu, The geometry of kernelized spectral clustering, Ann. Statist., 43 (2015), pp. 819-846.**

Some questions for further related research

- Stronger results for convergence rates of graph Laplacians.
- Explore other graph constructions : e.g.

$$W_{ij} = \frac{\tilde{W}_{ij}}{\tilde{d}_i^{\alpha} \tilde{d}_j^{\alpha}}$$

  Lafon & Coiffman (2008) Diffusion maps.
- PDE and calculus of variations ideas for Graph-based learning: a rapidly growing field ...

- N. Garcia Trillos and D. Slepcev, Continuum limit of total variation on point clouds, Archive for Rational Mechanics and Analysis, (2015), pp. 1-49
- E. Davis, S. Sethuraman, Consistency of modularity clustering on random geometric graphs. AAP. 2017.
- B. Oosting, T. Reeb, Consistency of Dirichlet Partitions. SIMA. 2017
- N. Garciaa Trillos and R. Murray, A new analytical approach to consistency and over
  tting in regularized empirical risk minimization. EJAM, 2017.
- D. Slepcev and M. Thorpe, Analysis of p-Laplacian Regularization in Semi-Supervised Learning. Preprint, 2017.
- J. Calder, The game theoretic p-Laplacian and semi-supervised learning with few labels. Preprint, 2017.
- N. Garcia Trillos and D. Sanz-Alonso, Continuum limit of posteriors in graph bayesian inverse prob lems. Preprint, 2017.

- N. Garcia Trillos, Z. Kaplan, T. Samakhoana, and D. Sanz-Alonso, On the consistency of graph-based Bayesian learning and the scalability of sampling algorithms. Preprint, 2017.
- N. Garcia Trillos, Gromov-Hausdorff limit of Wasserstein spaces on point clouds. Preprint 2017.
- T. Muller, M. D. Penrose, Optimal Cheeger cuts and bisections of random geometric graphs, Preprint (2018).
- M. Dunlop, D. Slepcev, A. Stuart, M. Thorpe, Large data and zero noise limits of graph-based semi supervised learning algorithms, Preprint (2018).

Thank you for your attention!