Consistency of objective functionals in semi-supervised learning

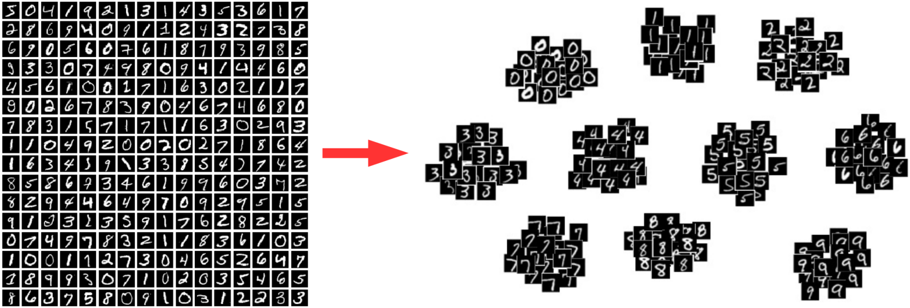
Dejan Slepčev
Carnegie Mellon University

Casa Matemática Oaxaca
May 1, 2017.

References

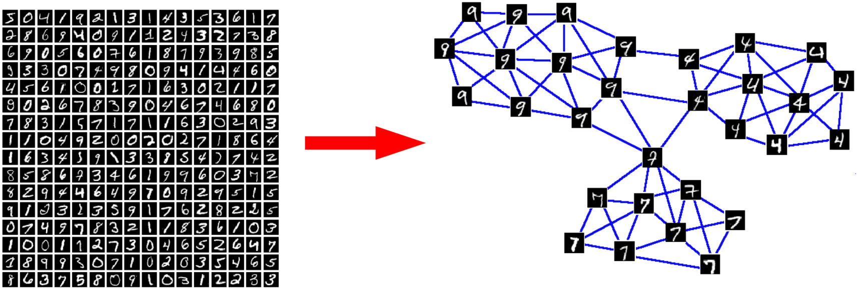
- Dunlop, S., Stuart, Thorpe, *Consistency of objective functionals in semi-supervised learning* in preparation.
- S.,Thorpe, *Consistency of p-Laplacian regularizations in semi-supervised learning* in preparation.
- García Trillos, Gerlach, Hein, and S., *Error bounds for spectral convergence of empirical graph Laplacians* in preparation.
- García Trillos and S., *On the rate of convergence of empirical measures in ∞-transportation distance*, Canad. J. Math, 67, (2015), pp. 1358-1383.
- García Trillos and S., *Continuum limit of total variation on point clouds*, Arch. Ration. Mech. Anal., 220 no. 1, (2016) 193-241.
- García Trillos, S., J. von Brecht, T. Laurent, and X. Bresson, *Consistency of Cheeger and ratio graph cuts*, J. Mach. Learn. Res. 17 (2016) 1-46.
- García Trillos, S., *A variational approach to the consistency of spectral clustering*, published online Applied and Computational Harmonic Analysis.

Clustering



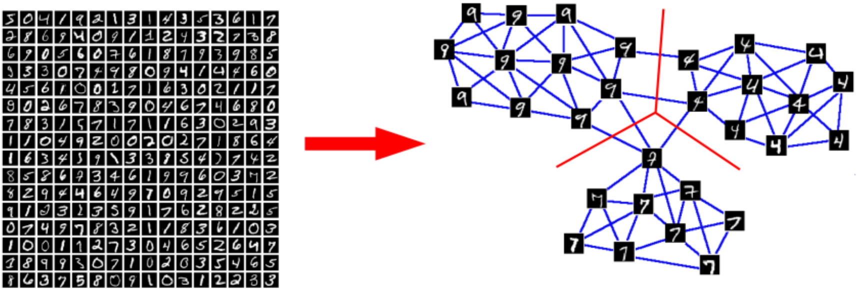
- Partition the data into meaningful groups.

Graph-Based Clustering



- Determine a similarity measure between images
- Construct a graph based on the similarity measure.

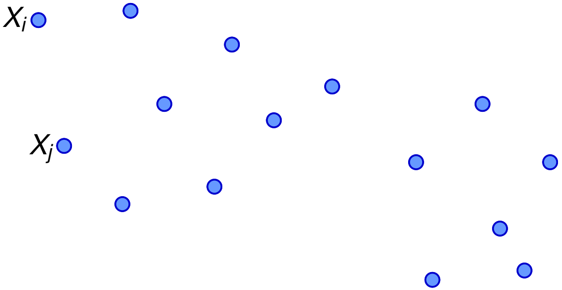
Graph-Based Clustering



- Determine a similarity measure between images
- Construct a graph based on the similarity measure.
- Partition the graph

From point clouds to graphs

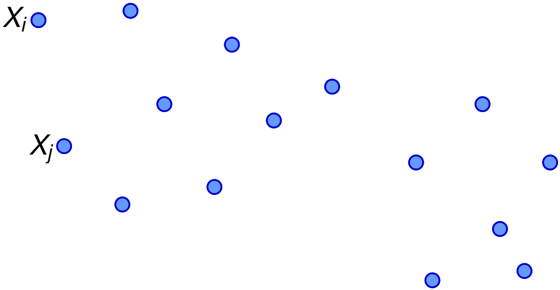
- Let $V = \{X_1, \ldots, X_n\}$ be a point cloud in $\mathbb{R}^d$:



- Connect nearby vertices: Edge weights $W_{i,j}$.

From point clouds to graphs

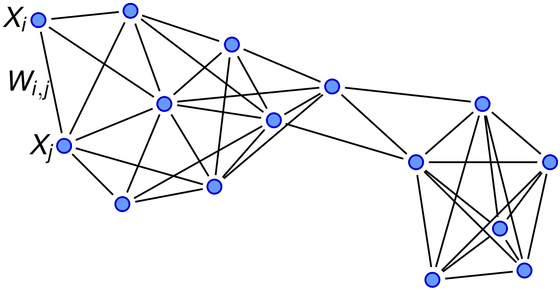
- Let $V = \{X_1, \ldots, X_n\}$ be a point cloud in $\mathbb{R}^d$:



- Connect nearby vertices: Edge weights $W_{i,j}$.

From point clouds to graphs

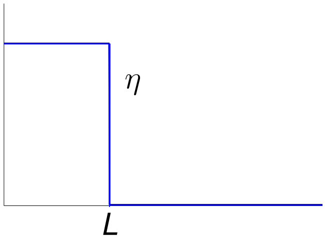
- Let $V = \{X_1, \ldots, X_n\}$ be a point cloud in $\mathbb{R}^d$:

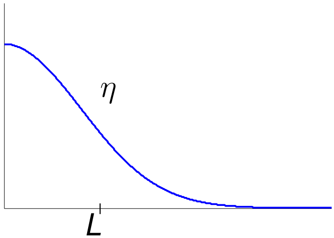


- Connect nearby vertices: Edge weights $W_{i,j}$.

- proximity based graphs

$$W_{i,j} = \eta(X_i - X_j)$$





- kNN graphs: Connect each vertex with its k nearest neighbors

k-means clustering

Given $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ find a set of k points $A = \{a_1, \ldots, a_k\}$ which minimizes

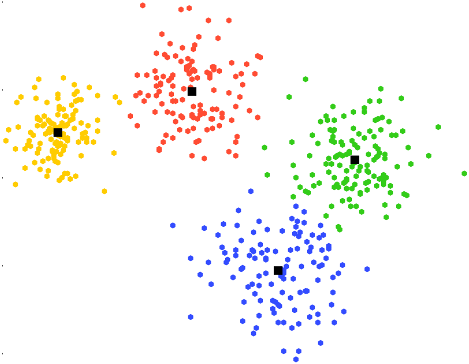
$$\min_A \frac{1}{n} \sum_{i=1}^{n} \mathrm{dist}(x_i, A)^2$$

where $\mathrm{dist}(x, A) = \min_{a \in A} |x - a|$.

Given $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ find a set of k points $A = \{a_1, \ldots, a_k\}$ which minimizes

$$\min_A \frac{1}{n} \sum_{i=1}^{n} \mathrm{dist}(x_i, A)^2$$

where $\mathrm{dist}(x, A) = \min_{a \in A} |x - a|$.



Spectral Clustering

Shi, Malik, '00, Ng, Jordan, Weiss, '01, Belkin, Niyogi, '01, von Luxburg '07

- $V_n = \{X_1, \dots, X_n\}$, similarity matrix W:

$$W_{ij} := \eta\left(|X_i - X_j|\right).$$

  The weighted degree of a vertex is $d_i = \sum_j W_{i,j}$.

- Dirichlet energy of $u_n : V_n \to \mathbb{R}$ is

$$F(u) = \frac{1}{2} \sum_{i,j} W_{ij} |u_n(X_i) - u_n(X_j)|^2.$$

- Associated operator is the (unnormalized) graph laplacian

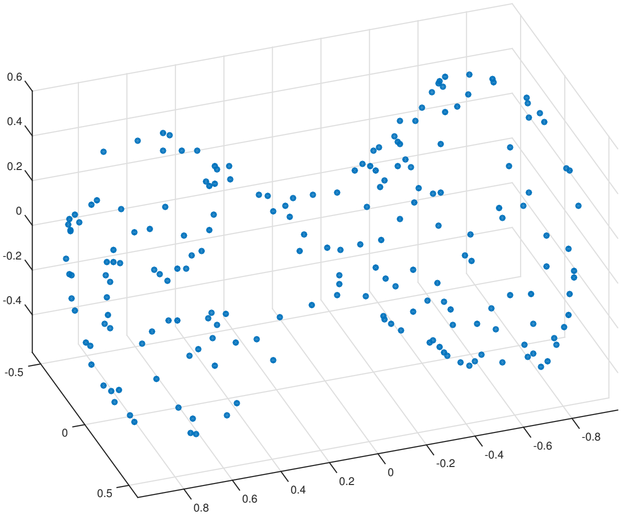
$$L = D - W,$$

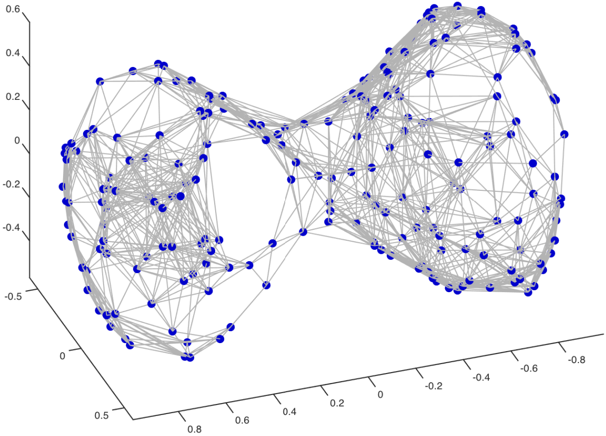
  where $D = \mathrm{diag}(d_1, \dots, d_n)$.

Input: Number of clusters k and similarity matrix W.

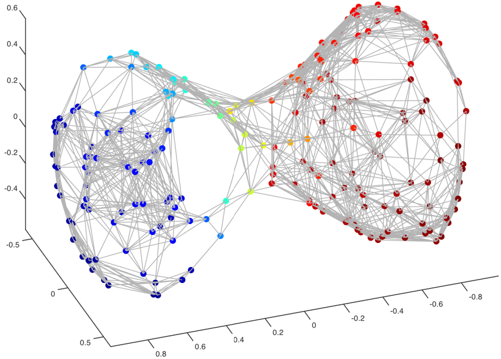
- Construct the unnormalized graph Laplacian L.
- Compute the eigenvectors $u_1, \ldots, u_k$ of L associated to the k smallest eigenvalues of L.
- Map the data into R^k: $x_i \mapsto (u_1(x_i), \ldots u_k(x_i)) =: y_i$
- Use the k-means algorithm to partition the set of points $\{y_1, \ldots, y_n\}$ into k groups, that we denote by $G_1, \ldots, G_k$.

Output: Clusters $G_1, \ldots, G_k$.

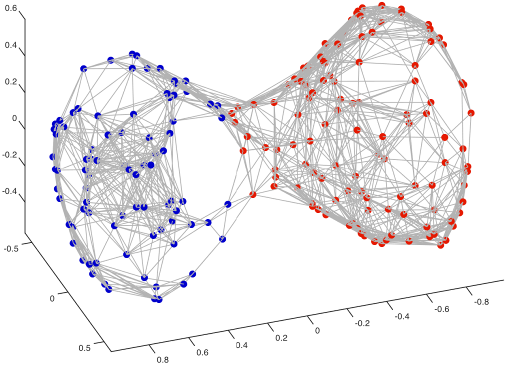

0.6
0.4
0.2
0
-0.2
-0.4
-0.5
0
0.5
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8


0.6
0.4
0.2
0
-0.2
-0.4
-0.5
0
0.5
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8

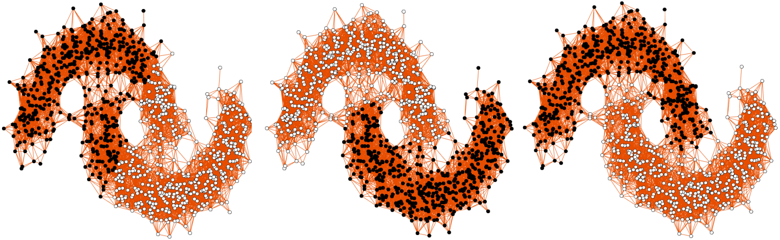
Spectral Clustering: an example



Spectral Clustering: an example



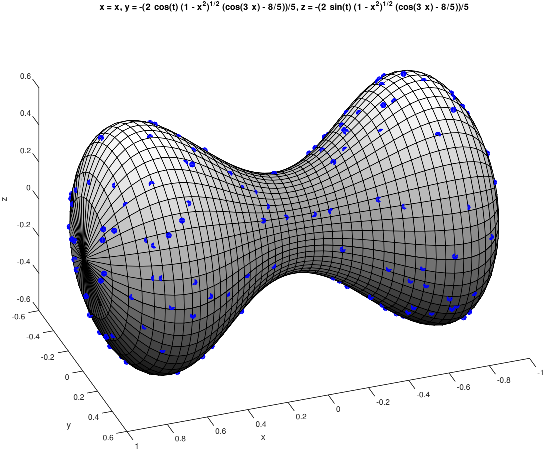
Comparison of Clustering Algorithms



(a) k - means (b) spectral (c) Cheeger cut

Ground Truth Assumption

Assume points $X_1, X_2, \ldots,$ are drawn i.i.d out of measure $d\nu = \rho d\,\mathrm{Vol}_{\mathcal{M}}$, where $\mathcal{M}$ is a compact manifold without boundary, and $0 < \rho < C$ is continuous.



Questions

Consistency of spectral clustering and graph Laplacians: *von Luxburg, Belkin, Bousquet '08, Belkin-Nyogi '07, Ting, Huang, Jordan '10, Singer, Wu '13, Burago, Ivanov, Kurylev '14, Shi, Sun '15*

- Does spectral clustering converge as $n \to \infty$?
- How should the connection distance be scaled as $n \to \infty$?
- What do the clusters converge to?
- Does the graph laplacian converge spectrally?
- Can one estimate the errors and obtain rates of convergence?

Spectral Clustering

- $V_n = \{X_1, \dots, X_n\}$, similarity matrix W:
$$W_{ij} := \frac{1}{\varepsilon^{d+2}} \eta \left(\frac{|X_i - X_j|}{\varepsilon} \right).$$
The weighted degree of a vertex is $d_i = \sum_j W_{i,j}$.
- Dirichlet energy of $u_n : V_n \to \mathbb{R}$ is
$$F(u) = \frac{1}{2} \sum_{i,j} W_{ij} |u_n(X_i) - u_n(X_j)|^2.$$
- Associated operator is the graph laplacian $L_n = D - W$, where $D = \mathrm{diag}(d_1, \dots, d_n)$.
- Spectrum has a variational characterization: The eigenvector corresponding to the second eigenvalue:
$$u_n := \arg\min \left\{ \sum_{i,j} W_{ij} |u(X_i) - u(X_j)|^2 \ : \ \sum_i u(X_i) = 0, \ \|u\|_2 = 1 \right\}$$

Consistency in Euclidean setting

Measure μ that data are sampled from is supported in $\overline{D}$ where D is bounded open set in $\mathbb{R}^d$ with Lipschitz boundary and the measure μ has continuous density ρ on D such that $\alpha < \rho < \frac{1}{\alpha}$ on D, for some $\alpha > 0$.

The spectral limit of the unweighted graph laplacian is given by the following eigenvalue problem.

$$\begin{aligned} L_c u = -\frac{1}{\rho}\operatorname{div}(\rho^2 \nabla u) = \lambda_2 u \quad & \text{in } D \\ \frac{\partial u}{\partial n} = 0 \qquad & \text{on } \partial D. \end{aligned}$$

The operator L_c describing the equation is self-adjoint with respect to the ρ-weighted L^2 inner product on D:

$$\langle u, v \rangle = \int_D u(x)v(x)\rho(x)dx$$

Consistency of Spectral Clustering in $\mathbb{R}^d$

Theorem (García Trillos and S., ACHA '16)

Assume $h \to 0$ as $n \to \infty$ and

$$\varepsilon^d \gg \begin{cases} \frac{(\ln n)^{\frac{3}{2}}}{n} & \text{if } d = 2 \\ \frac{\ln n}{n} & \text{if } d \geq 3 \end{cases}$$

Then

(i) eigenvalues of the graph laplacian converge to eigenvalues of L_c

(ii) eigenvectors of the graph laplacian converge (along a subsequence) to eigenfunctions of L_c.

(iii) the clusters obtained by spectral clustering converge to clustering obtained by spectral clustering in continuum setting.

Comment of ε_n

- We require

$$\varepsilon_n \gg \frac{(\log n)^{3/4}}{n^{1/2}} \quad \text{if } d = 2$$
$$\varepsilon_n \gg \frac{(\log n)^{1/d}}{n^{1/d}} \quad \text{if } d \geq 3.$$

- Note that for $d \geq 3$ this means that typical degree $\gg \log(n)$.
- Does convergence hold if fewer than $\log(n)$ neighbors are connected to?

Comment of ε_n

- We require

$$\varepsilon_n \gg \frac{(\log n)^{3/4}}{n^{1/2}} \quad \text{if } d = 2$$
$$\varepsilon_n \gg \frac{(\log n)^{1/d}}{n^{1/d}} \quad \text{if } d \geq 3.$$

- Note that for $d \geq 3$ this means that typical degree $\gg \log(n)$.
- Does convergence hold if fewer than $\log(n)$ neighbors are connected to?

  No. There exists $c > 0$ such that $\varepsilon_n < c\frac{\log(n)^{1/d}}{n^{1/d}}$ then with probability one the random geometric graph is asymptotically disconnected. This implies that for large enough n, $\min GC_{n,\varepsilon_n} = 0$. While $\inf C > 0$.

  So for $d \geq 3$ the condition is optimal in terms of scaling.

Optimal Transportation for $p = \infty$

$\infty-$transportation distance:

$$d_\infty(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \text{esssup}_\pi \{|x - y| \;:\; x \in X, y \in Y\}$$

- If $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ and $\nu = \frac{1}{n}\sum_{j=1}^n \delta_{y_j}$ then

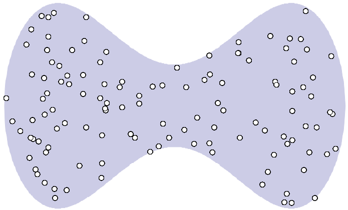
$$d_\infty(\mu, \nu) = \min_{\sigma-\text{permutation}} \max_i |x_i - y_{\sigma(i)}|.$$

- If μ has density then OT map, T exists (Champion, De Pascale, Juutinen 2008) and

$$d_\infty(\mu, \nu) = \|T(x) - x\|_{L^\infty(\mu)}.$$

∞-OT between a measure and its random sample

Optimal matchings in dimension $\mathbf{d} \geq \mathbf{3}$: *Ajtai-Komlós-Tusnády (1983), Yukich and Shor (1991), Garcia Trillos and S. (2014)*



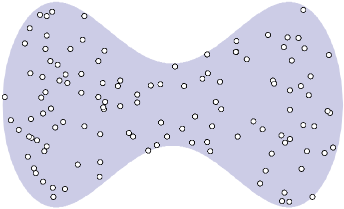
Theorem

There are constants $c > 0$ and $C > 0$ (depending on d) such that with probability one we can find a sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$ from ν_0 to ν_n ($T_{n\#}\nu_0 = \nu_n$) and such that:

$$c \leq \liminf_{n\to\infty} \frac{n^{1/d} \|Id - T_n\|_\infty}{(\log n)^{1/d}} \leq \limsup_{n\to\infty} \frac{n^{1/d} \|Id - T_n\|_\infty}{(\log n)^{1/d}} \leq C.$$

∞-OT between a measure and its random sample

Optimal matchings in dimension $\mathbf{d} = \mathbf{2}$: *Leighton and Shor (1986), new proof by Talagrand (2005), Garcia Trillos and S. (2014)*



Theorem

There are constants $c > 0$ and $C > 0$ such that with probability one we can find a sequence of transportation maps $\{T_n\}_{n \in \mathbb{N}}$ from ν_0 to ν_n ($T_{n\#}\nu_0 = \nu_n$) and such that:

$$c \leq \liminf_{n \to \infty} \frac{n^{1/2} \| Id - T_n \|_\infty}{(\log n)^{3/4}} \leq \limsup_{n \to \infty} \frac{n^{1/2} \| Id - T_n \|_\infty}{(\log n)^{3/4}} \leq C. \tag{1}$$

work in progress with García Trillos, Gerlach, and Hein. Relies on work by Burago, Ivanov Kurylev.

$\mathcal{M}$ compact manifold of dimension m.
The measure μ on $\mathcal{M}$ the data are sampled from has density ρ with respect to volume form on $\mathcal{M}$, such that $\alpha \leq \rho \leq \frac{1}{\alpha}$ for some $\alpha > 0$ and ρ is Lipschitz continuous.
The continuum operator is a weighted Laplace-Beltrami operator

$$u \mapsto \frac{1}{\rho} \operatorname{div}_{\mathcal{M}}(\rho^2 \operatorname{grad} u).$$

This operator is symmetric with respect to $L^2(d\mu)$:

$$\|u\|^2_{L^2(d\mu)} = \int_{\mathcal{M}} u^2 d\mu.$$

It has a spectrum

$$0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots .$$

with corresponding orthornomal set of eigenfunctions u_k, $k = 1, \ldots$.

Transportation estimates

Let $\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ be the empirical measure of the random i.i.d sample.

Theorem

For any $\beta > 1$ and every $n \in \mathbb{N}$ there exist a transportation map $T_n \colon \mathcal{M} \to X$ and a constant A such that

$$\ell = \sup_{x \in \mathcal{M}} d(x, T_n(x)) \leq A \begin{cases} \frac{\log(n)^{3/4}}{n^{1/2}}, & \text{if } m = 2, \\ \frac{(\log n)^{1/m}}{n^{1/m}}, & \text{if } m \geq 3, \end{cases}$$

holds with probability at least $1 - C_{K,\mathrm{Vol}(\mathcal{M}),m,i_0} \cdot n^{-\beta}$, where A depends only on K, i_0, R, m, $\mathrm{Vol}(\mathcal{M})$, α and β.

- K – upper bound on absolute value of sectional curvature
- i_0 – injectivity radius
- R – reach of $\mathcal{M}$ is $\mathbb{R}^d$

Consistency of Spectral Clustering in Manifold Setting

Theorem (García Trillos, Gerlach, Hein and S.)

With high probability, for every $k \in \{1, \dots, n\}$ there exists a constant $C > 0$ depending on K, R, m, p, ρ, $\vec{m}$, η, and $\lambda_k(\mathcal{M})$ such that

$$|\lambda_k(\Gamma) - \lambda_k(\mathcal{M})| \leq C\left(\varepsilon + \frac{\ell}{\varepsilon}\right),$$

whenever $\ell < h \ll 1$.

- ε – averaging length scale
- ℓ – transportation length scale
- K – upper bound on absolute value of sectional curvature
- R – reach of $\mathcal{M}$ is $\mathbb{R}^d$

Consistency of Spectral Clustering in Manifold Setting

- ε – averaging length scale
- ℓ – transportation length scale

Theorem (García Trillos, Gerlach, Hein and S.)

With high probability, for every $k \in \{1, \ldots, n\}$ there exists a constant $C > 0$ depending on K, R, m, p, ρ, $\vec{m}$, η, and $\lambda_k(\mathcal{M})$ such that

$$\|u_k^n - u_k\|_{L^2} \leq C\left(\varepsilon + \frac{\ell}{\varepsilon}\right),$$

whenever $\ell < h \ll 1$.

where $u_k^n : V_n \to \mathbb{R}$, $u : \mathcal{M} \to \mathbb{R}$, and

$$L_n u_k^n = \lambda_k(\Gamma_n) u_n^k$$
$$L_c u_k = \lambda_k(\mathcal{M}) u_k.$$

Consistency of Spectral Clustering in Manifold Setting

- ε – averaging length scale
- ℓ – transportation length scale

Theorem (García Trillos, Gerlach, Hein and S.)

With high probability, for every $k \in \{1, \dots, n\}$ there exists a constant $C > 0$ depending on K, R, m, p, ρ, $\vec{m}$, η, and $\lambda_k(\mathcal{M})$ such that

$$d_{TL^2}((\mu_n, u_k^n), (\mu, u_k)) \leq C\left(\varepsilon + \frac{\ell}{\varepsilon}\right),$$

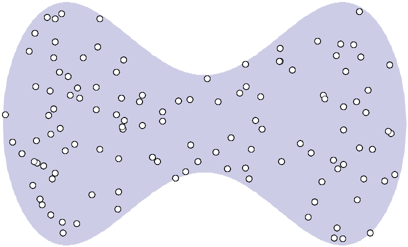
whenever $\ell < h \ll 1$.

where $u_k^n : V_n \to \mathbb{R}$, $u : \mathcal{M} \to \mathbb{R}$, and

$$L_n u_k^n = \lambda_k(\Gamma_n) u_n^k$$
$$L_c u_k = \lambda_k(\mathcal{M}) u_k.$$

Topology

Consider domain D and $V_n = \{X_1, \ldots, X_n\}$ random i.i.d points.

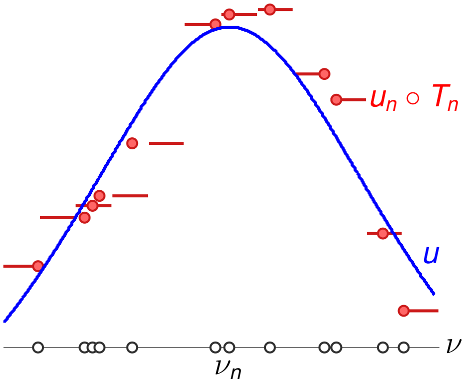


- How to compare $u_n : V_n \to \mathbb{R}$ and $u : D \to \mathbb{R}$ in a way consistent with L^1 topology?

Note that $u \in L^1(\nu)$ and $u_n \in L^1(\nu_n)$, where $\nu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$.

Topology

Consider domain D and $V_n = \{X_1, \ldots, X_n\}$ random i.i.d points.



- Let T_n be a transportation map from ν to ν_n.

Let ν be a measure with density ρ, supported on the domain D.

We need to compare values at nearby points. Thus we also penalize transport $|T_n(x) - x|$.

Metric

For $u \in L^1(\nu)$ and $u_n \in L^1(\nu_n)$

$$d((\nu, u), (\nu_n, u_n)) = \inf_{T_{n\sharp}\nu = \nu_n} \int_D \left(|u_n(T_n(x)) - u(x)| + |T_n(x) - x|\right) \rho(x) dx$$

where

$$T_{n\sharp}\nu = \nu_n$$

TL^p Space

Definition

$$TL^p = \{(\nu, f) \ : \ \nu \in \mathcal{P}(D), \ f \in L^p(\nu)\}$$

$$d^p_{TL^p}((\nu, f), (\sigma, g)) = \inf_{\pi \in \Pi(\nu,\sigma)} \int_{D\times D} |y - x|^p + |g(y) - f(x))|^p d\pi(x, y).$$

where

$$\Pi(\nu, \sigma) = \{\pi \in \mathcal{P}(D \times D) \ : \ \pi(A \times D) = \nu(A), \ \pi(D \times A) = \sigma(A)\}.$$

Lemma

(TL^p, d_{TL^p}) is a metric space.

The topology of TL^p agrees with the L^p convergence in the sense that

- $(\nu, f_n) \xrightarrow{TL^p} (\nu, f)$ iff $f_n \xrightarrow{L^p(\nu)} f$

- $(\nu, f_n) \xrightarrow{TL^p} (\nu, f)$ iff $f_n \xrightarrow{L^p(\nu)} f$
- $(\nu_n, f_n) \xrightarrow{TL^p} (\nu, f)$ iff the measures $(I \times f_n)_\sharp \nu_n$ weakly converge to $(I \times f)_\sharp \nu$. That is if graphs, considered as measures converge weakly.
- The space TL^p is not complete. Its completion are the probability measures on the product space $D \times \mathbb{R}$.

If $(\nu_n, f_n) \xrightarrow{TL^p} (\nu, f)$ then there exists a sequence of transportation plans ν_n such that

$$\int_{D\times D} |x-y|^p d\pi_n(x,y) \longrightarrow 0 \quad \text{as } n \to \infty. \tag{2}$$

We call a sequence of transportation plans $\pi_n \in \Pi(\nu_n, \nu)$ **stagnating** if it satisfies (2).

Stagnating sequence: $\int_{D\times D}|x-y|^p d\pi_n(x,y) \longrightarrow 0$

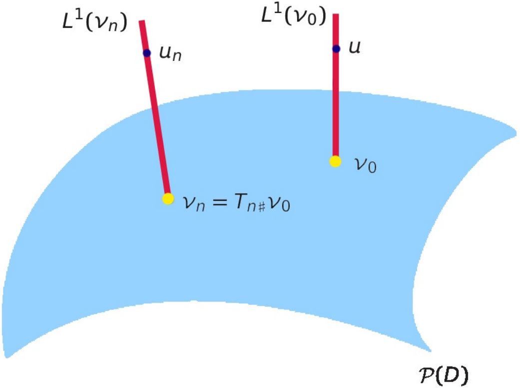
TFAE:

1. $(\nu_n, f_n) \xrightarrow{TL^p} (\nu, f)$ as $n \to \infty$.
2. $\nu_n \rightharpoonup \nu$ and **there exists** a stagnating sequence of transportation plans $\{\pi_n\}_{n\in\mathbb{N}}$ for which

$$\iint_{D\times D} |f(x) - f_n(y)|^p \, d\pi_n(x,y) \to 0, \text{ as } n \to \infty. \tag{3}$$

3. $\nu_n \rightharpoonup \nu$ and **for every** stagnating sequence of transportation plans π_n, (3) holds.

Formally $TL^p(D)$ is a fiber bundle over $\mathcal{P}(D)$.



Lemma

Let $p \geq 1$ and let $\{\nu_n\}_{n\in\mathbb{N}}$ and ν be Borel probability measures on $\mathbb{R}^d$ with finite second moments. Let $F_n \in L^p(\nu_n, \mathbb{R}^d, \mathbb{R}^k)$ and $F \in L^p(\nu, \mathbb{R}^d, \mathbb{R}^k)$. Consider the measures $\tilde{\nu}_n, = F_{n\sharp}\nu_n$ and $\tilde{\nu}, = F_\sharp \nu$. Finally, let $\tilde{f}_n \in L^p(\tilde{\nu}_n, \mathbb{R}^k, \mathbb{R})$ and $\tilde{f} \in L^p(\tilde{\nu}, \mathbb{R}^k, \mathbb{R})$. If

$$(\nu_n, F_n) \xrightarrow{TL^p} (\nu, F) \quad \text{as } n \to \infty,$$

and

$$(\tilde{\nu}_n, \tilde{f}_n) \xrightarrow{TL^p} (\tilde{\nu}, \tilde{f}) \quad \text{as } n \to \infty.$$

Then,

$$(\nu_n, \tilde{f}_n \circ F_n) \xrightarrow{TL^p} (\nu, \tilde{f} \circ F_n) \quad \text{as } n \to \infty.$$

Consistency of Spectral Clustering in manifold setting

Theorem (García Trillos and S., ACHA '16)

Assume $h \to 0$ as $n \to \infty$ and

$$\varepsilon^d \gg \begin{cases} \frac{(\ln n)^{\frac{3}{2}}}{n} & \text{if } d \geq 2 \\ \frac{\ln n}{n} & \text{if } d \geq 3 \end{cases}$$

Then

(i) eigenvalues of the graph laplacian converge to eigenvalues of L_c

(ii) eigenvectors of the graph laplacian converge (along a subsequence) to eigenfunctions of L_c.

(iii) the clusters obtained by spectral clustering converge to clustering obtained by spectral clustering in continuum setting.

Functionals in semi-supervised learning

x_i- points, y_i- real valued labels
Assume we are given k labeled points

$$(x_1, y_1), \dots (x_k, y_k)$$

and a random sample $x_{k+1}, \dots x_n$.

Q. How to label the rest of the points?

Zhu, Ghahramani, and Lafferty '03 proposed the following

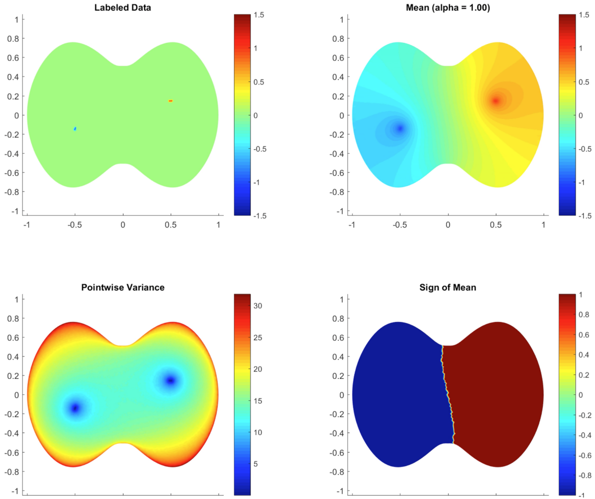
Harmonic SSL

Minimize $$E(u) = \frac{1}{n^2} \sum_{i,j} W_{i,j}(u_i - u_j)^2$$

subject to constraint $$u_i = y_i \quad \text{for } i = 1, \dots, k.$$

Harmonic semi-supervised learning

Nadler, Srebro, and Zhou '09 observed that solutions are spiky as $n \to \infty$, while the Dirichlet energy is decreasing. [Also see Wahba '90.]



Harmonic semi-supervised learning

It can be shown that if $W_{i,j} = \frac{1}{\varepsilon_n^2}\eta_{\varepsilon_n}(x_i - x_j)$ and

$$\varepsilon_n^d \gg \begin{cases} \frac{(\ln n)^{\frac{3}{2}}}{n} & \text{if } d = 2 \\ \frac{\ln n}{n} & \text{if } d \geq 3 \end{cases}$$

then the minimizers u^n of

$$E(u^n) = \frac{1}{n^2}\sum_{i,j} W_{i,j}(u_i^n - u_j^n)^2$$

subject to constraint $$u_i^n = y_i \quad \text{for } i = 1, \dots, k.$$

converge along a subsequence to a “harmonic” function which in general does not respect the labels.

Harmonic semi-supervised learning II

$x_1, \dots x_n$ random sample of a measure μ with density ρ on Ω.
Points in subdomain $\Omega^+ \subset \Omega$ are labeled: $y_i = f(x_i)$ for $x_i \in \Omega^+$.

Consider, as did *Bertozzi, Luo, Stuart, Zygalakis*, minimizing

Harmonic SSL

$$E(u^n) = \frac{1}{n^2} \sum_{i,j} W_{i,j}(u_i^n - u_j^n)^2 + \frac{1}{\gamma^2} \frac{1}{n} \sum_{i\,:\,x_i \in \Omega^+} |u_i^n - f(x_i)|^2$$

Theorem (Dunlop, Stuart, S. Thorpe)

Under standard assumptions the minimizers u^n converge in TL^2 to the minimizer of

$$E(u) = \sigma \int_\Omega |\nabla u|^2 \rho^2 dx + \frac{1}{\gamma^2} \int_{\Omega^+} |u(x) - f(x)|^2 \rho(x) dx$$

Higher order regularizations

Related work by *Zhou, Belkin '11*.
Given are k labeled points, $(x_1, y_1), \dots (x_k, y_k)$, and a random sample $x_{k+1}, \dots x_n$.

Using graph laplacian L_n we define

$$A_n = (L_n + \tau^2 I)^\alpha.$$

Power of a symmetric matrix is defined by $A^\alpha = PD^\alpha P^{-1}$ for $A = PDP^{-1}$.

Higher order SSL

Minimize	$E(u) = \frac{1}{2}\langle u^n, A_n u^n\rangle_{\mu_n}$
subject to constraint	$u_i^n = y_i \quad \text{for } i = 1, \dots, k.$

Higher order regularizations

$$A_n = (L_n + \tau^2 I)^\alpha.$$

Higher order SSL

$$\begin{aligned} &\text{Minimize} && E(u) = \frac{1}{2}\langle u^n, A_n u^n\rangle_{\mu_n} \\ &\text{subject to constraint} && u_i^n = y_i \quad \text{for } i = 1, \dots, k. \end{aligned}$$

Theorem (Dunlop, Stuart, S. Thorpe)

For $\alpha > \frac{d}{2}$, under usual assumptions, minimizers u^n converge in TL^2 to the

$$\begin{aligned} &\text{minimizer of} && E(u) = \sigma \int_\Omega u(x)(Au)(x)\rho(x)dx \\ &\text{subject to constraint} && u(x_i) = y_i \quad \text{for } i = 1, \dots, k. \end{aligned}$$

where $A = (\sigma L_c + \tau I)^\alpha$ and $L_c u = -\frac{1}{\rho}\operatorname{div}(\rho^2 \nabla u)$.