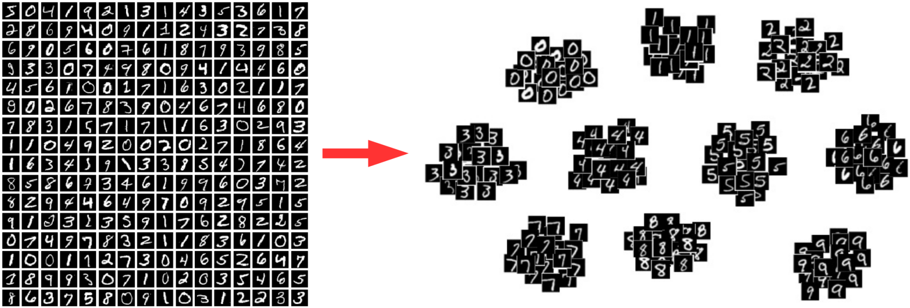
Variational problems on graphs and their continuum limits

Dejan Slepčev
Carnegie Mellon University

Calculus of Variations, Optimal Transportation, and Geometric Measure Theory: from Theory to Applications
Université Lyon 1
July 6, 2016.

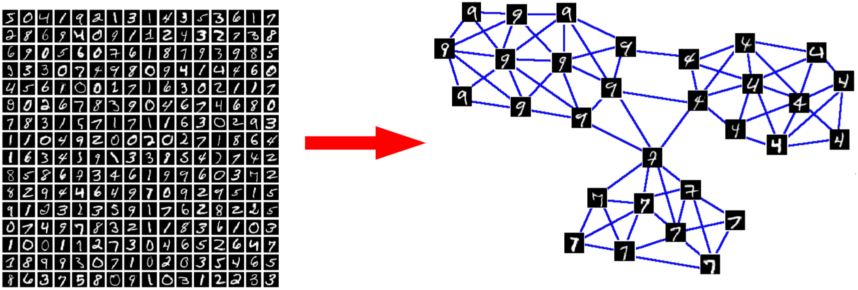
References

- 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*, accepted for publication in J. Mach. Learn. Res.
- García Trillos, S., *A variational approach to the consistency of spectral clustering*, preprint.
- García Trillos, S., J. von Brecht, *Estimating perimeter from graph cuts*, preprint.
- García Trillos, Gerlach, Hein, and S. *Error bounds for spectral convergence of empirical graph Laplacians* in preparation.



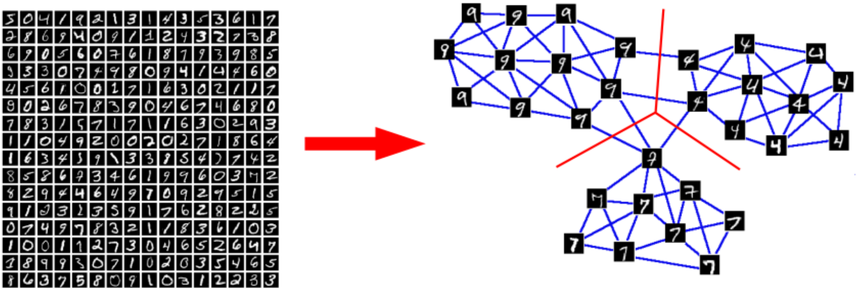
- Partition the data into meaningful groups.

Graph-Based Clustering



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

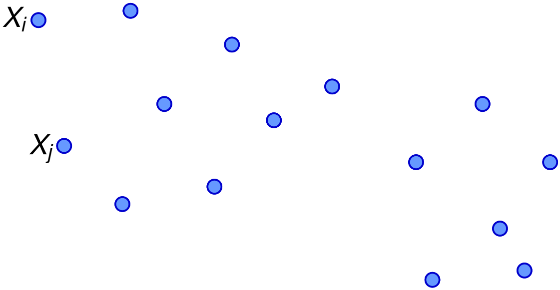
Graph-Based Clustering



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

From point clouds to graphs

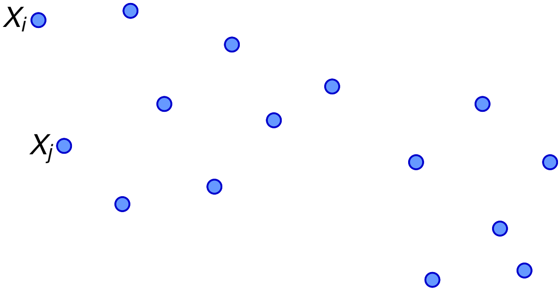
- Let $V = \{X_1, \ldots, X_n\}$ be a point cloud in $\mathbb{R}^d$:



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

From point clouds to graphs

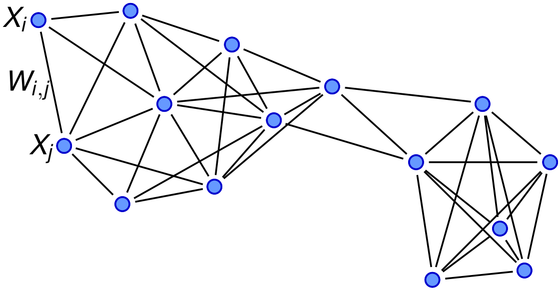
- Let $V = \{X_1, \ldots, X_n\}$ be a point cloud in $\mathbb{R}^d$:



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

From point clouds to graphs

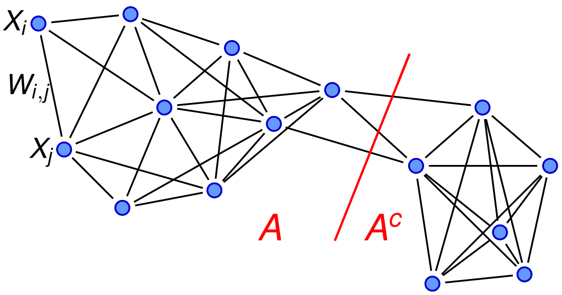
- Let $V = \{X_1, \ldots, X_n\}$ be a point cloud in $\mathbb{R}^d$:



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

Graph cut

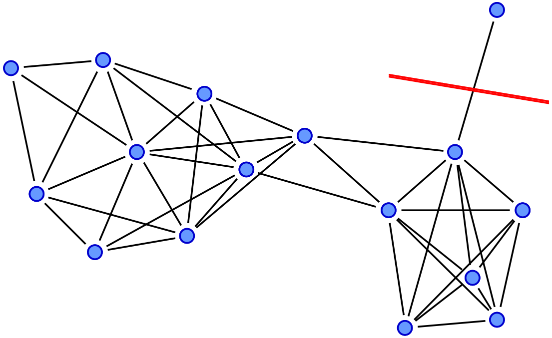
- Let $V = \{X_1, \ldots, X_n\}$ be a point cloud in $\mathbb{R}^d$:



- Connect nearby vertices: Edge weights $W_{i,j}$
- Graph Cut: $A \subset V$.

$$Cut(A, A^c) = \sum_{i \in A} \sum_{j \in A^c} W_{i,j}.$$

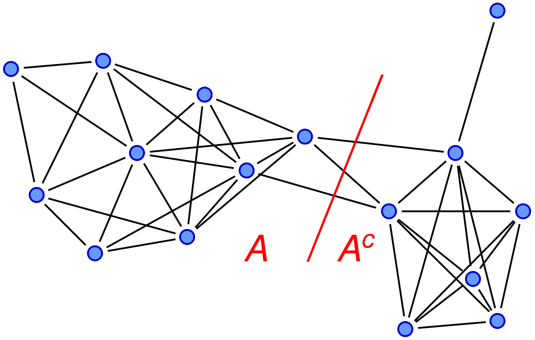
- Let $V = \{X_1, \ldots, X_n\}$ be a point cloud in $\mathbb{R}^d$:



- Connect nearby vertices: Edge weights $W_{i,j}$
- Minimize: $A \subset V$.

$$Cut(A, A^c) = \sum_{i \in A} \sum_{j \in A^c} W_{i,j}.$$

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



- Graph Cut: $A \subset V$.

$$Cut(A, A^c) = \sum_{i \in A} \sum_{j \in A^c} W_{i,j}.$$

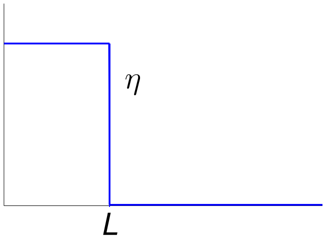
- Cheeger Cut: Minimize

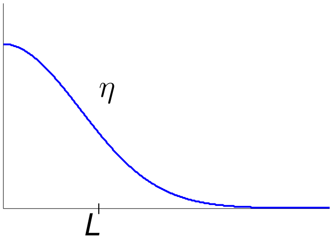
$$GC(A) = \frac{Cut(A, A^c)}{\min\{|A|, |A^c|\}}.$$

Graph Constructions

- proximity based graphs

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





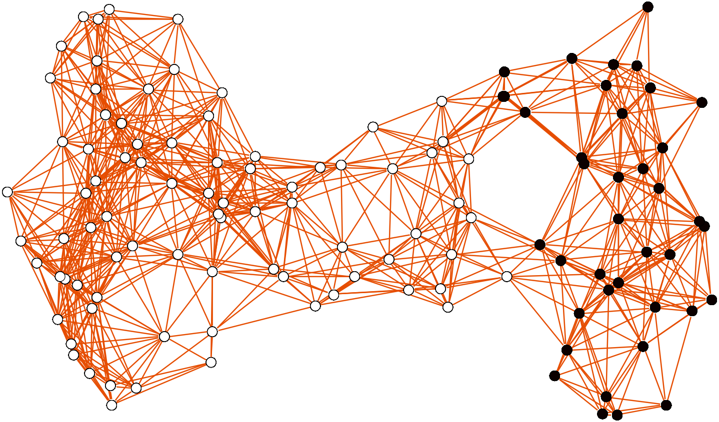
- kNN graphs: Connect each vertex with its k nearest neighbors

Task

Minimize $$GC(A) = \frac{\sum_{i \in A} \sum_{j \in A^c} W_{i,j}}{\min\{|A|, |A^c|\}}$$

Task

Minimize
$$GC(A) = \frac{\sum_{i \in A} \sum_{j \in A^c} W_{i,j}}{\min\{|A|, |A^c|\}}$$



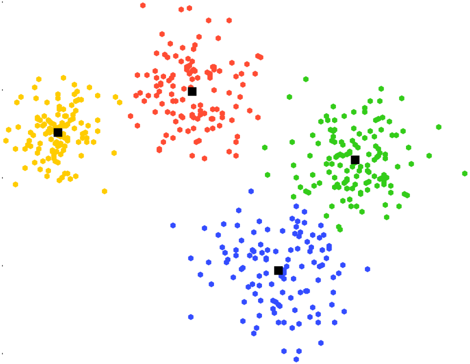
Algorithm of Bresson, Laurent, Uminsky and von Brecht (2013).

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



- $V_n = \{X_1, \ldots, X_n\}$, similarity matrix W, as before:

$$W_{ij} := \frac{1}{\varepsilon^d} \eta \left(|X_i - X_j| / \varepsilon \right).$$

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 (unnormalized) graph laplacian

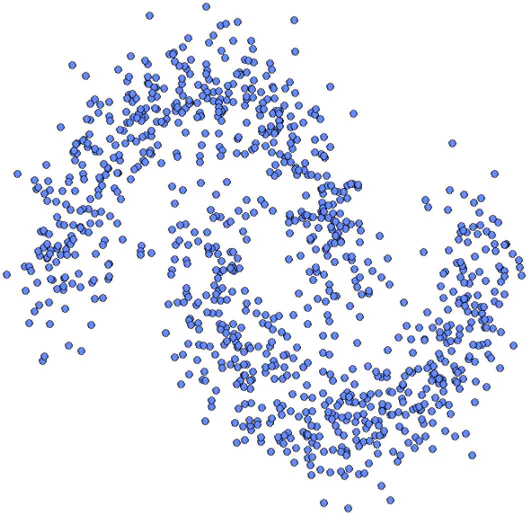
$$L = D - W,$$

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

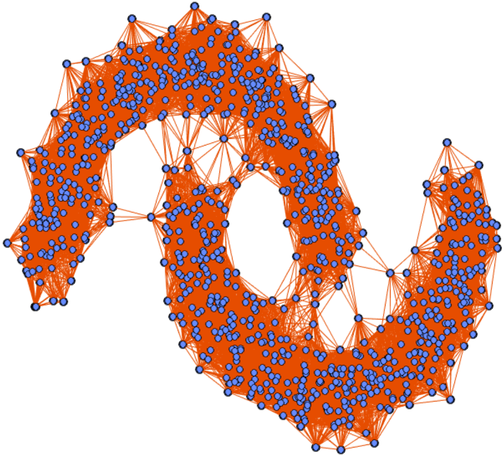
Input: Number of clusters k and similarity matrix W.

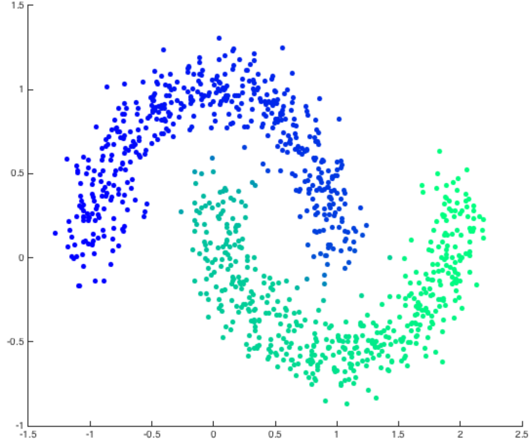
- Construct the unnormalized graph Laplacian L.
- Compute the eigenvectors $u_1, \ldots, u_k$ of L associated to the k smallest eigenvalues of L.
- Define the matrix $U \in \mathbb{R}^{k \times n}$, where the i-th row of U is the vector u_i.
- For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the i-th column of U.
- 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 $G_1, \ldots, G_k$.



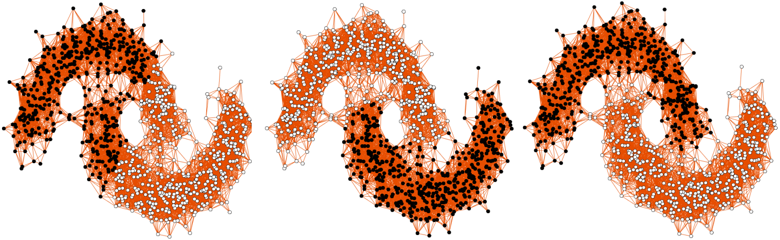
Spectral Clustering: Two moons (easy)





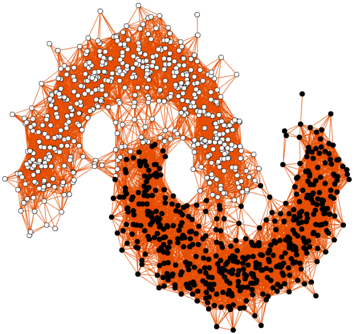
1D embedding: $x_i \mapsto v_2(x_i)$

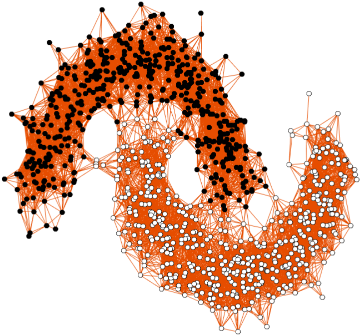
Comparison of Clustering Algorithms



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

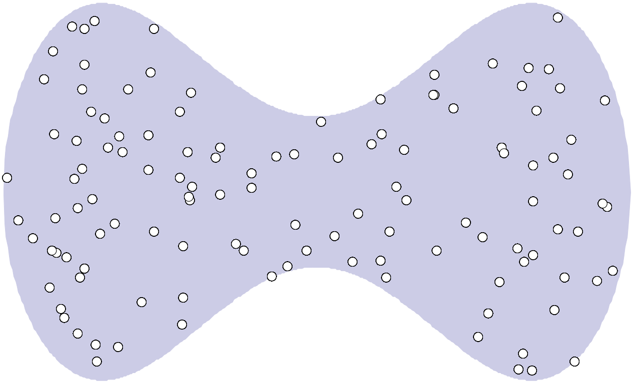
Comparison of Clustering Algorithms


(d) spectral


(e) Cheeger cut

Ground Truth Assumption

Assume points $X_1, X_2, \ldots,$ are drawn i.i.d out of measure $d\nu = \rho dx$



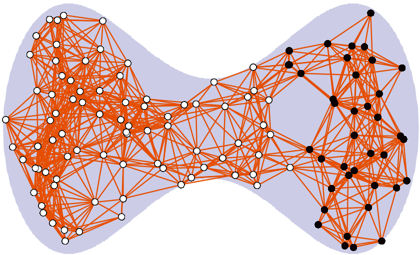
Consistency of Cheeger cut clustering

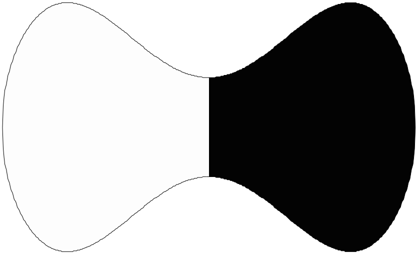
Consistency of clustering

Do the minimizers of

$$GC(A) = \frac{\sum_{i \in A} \sum_{j \in A^c} W_{i,j}}{\min\{|A|, |A^c|\}}$$

converge as the number of data points $n \to \infty$?
Can one characterize the limiting object as a minimizer of a continuum functional?





Graph Total Variation

Graph total variation

For a function $u : V \to \mathbb{R}$

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

where $u_i = u(X_i)$.

Note that for a set of vertices $A \subset V$

$$GTV_n(\chi_A) = \frac{1}{n^2} Cut(A, A^c)$$

where χ_A is the characteristic function of A

$$\chi_A(X_i) = \begin{cases} 1 & \text{if } x_i \in A \\ 0 & \text{otherwise.} \end{cases}$$

Relaxed Problem

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

Balance term

$$B_n(u) = \frac{1}{n} \min_{c \in \mathbb{R}} \sum_i |u_i - c|$$

Note that

$$B_n(\chi_A) = \frac{1}{n} \min\{|A|, |A^c|\}.$$

Relaxed problem

Minimize

$$GC_n(u) = \frac{GTV_n(u)}{B_n(u)}$$

Theorem

Relaxation is exact: There exists a set of vertices A_n such that $u_n = \chi_{A_n}$ minimizes GC_n.

Total variation in continuum setting

- $d\nu = \rho dx$ probability measure, $\mathrm{supp}(\nu) = D$, $0 < \lambda \leq \rho \leq \frac{1}{\lambda}$ on D.

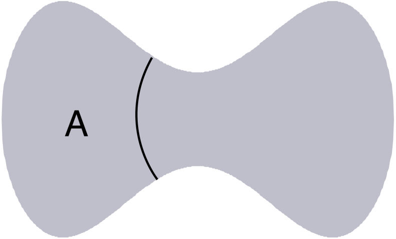
Weighted relative perimeter

Given $A \subset D$

$$P(A; D, \rho^2) = \int_{D \cap \partial A} \rho^2 dS_{d-1}$$

Weighted TV

$$TV(u, \rho^2) = \int_D |\nabla u| \rho^2 dx$$



Total variation in continuum setting

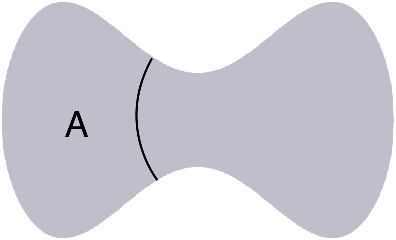
- $d\nu = \rho dx$ probability measure, $\mathrm{supp}(\nu) = D$, $0 < \lambda \leq \rho \leq \frac{1}{\lambda}$ on D.

Weighted relative perimeter

Given $A \subset D$
$$P(A; D, \rho^2) = \int_{D \cap \partial A} \rho^2 dS_{d-1} = TV(\chi_A, \rho^2)$$

Weighted TV

$$TV(u, \rho^2) = \sup \left\{ \int_D u \, \mathrm{div}(\phi) dx \; : \; |\phi| \leq \rho^2 \;\; , \; \phi \in C_c^\infty(D, \mathbb{R}^d) \right\}$$



Clustering in continuum setting

- ν probability measure with compact support $\mathrm{supp}(\nu) = D$.
- ν has continuous on D density ρ and $0 < \lambda \leq \rho \leq \frac{1}{\lambda}$ on D.

Weighted TV

$$TV(u, \rho^2) = \sup \left\{ \int_D u \,\mathrm{div}(\phi) dx \; : \; |\phi| \leq \rho^2 \;\; , \; \phi \in C_c^\infty(D, \mathbb{R}^d) \right\}$$

Weighted relative perimeter

Given $A \subset D$ $\qquad P(A; D, \rho^2) = TV(\chi_A, \rho^2)$

Balance term

$$B(A) = \min\{|A|, 1 - |A|\} \qquad \text{where } |A| = \nu(A).$$

Weighted Cheeger Cut: Minimize

$$C(A) = \frac{P(A; D, \rho^2)}{B(A)}$$

Relaxation in continuum setting

- ν probability measure with compact support $\mathrm{supp}(\nu) = D$.
- ν has continuous on D density ρ and $0 < \lambda \leq \rho \leq \frac{1}{\lambda}$ on D.

Weighted TV

$$TV(u, \rho^2) = \sup \left\{ \int_D u \, \mathrm{div}(\phi) dx \; : \; |\phi| \leq \rho^2 \; , \; \phi \in C_c^\infty(D, \mathbb{R}^d) \right\}$$

Balance term

$$B(u) = \min_{c \in \mathbb{R}} \int_D |u(x) - c| \rho(x) dx$$

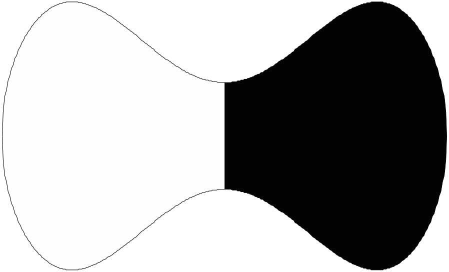
Minimize

$$C(u) = \frac{TV(u, \rho^2)}{B(u)}$$

Clustering in continuum setting

Minimize

$$C(u) = \frac{TV(u, \rho^2)}{B(u)}$$



Localizing the kernel as $n \to \infty$

$$\eta_\varepsilon(z) = \frac{1}{\varepsilon^d} \eta\left(\frac{z}{\varepsilon}\right).$$

Consistency of clustering II

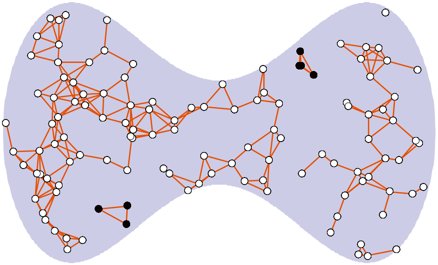
Do the minimizers of

$$GC_{n,\varepsilon_n}(u^n) = \frac{1}{n} \frac{\frac{1}{\varepsilon_n}\sum_{i,j} \eta_{\varepsilon_n}(X_i - X_j)\,|u_i^n - u_j^n|}{\min_{c\in\mathbb{R}} \sum_i |u_i^n - c|}$$

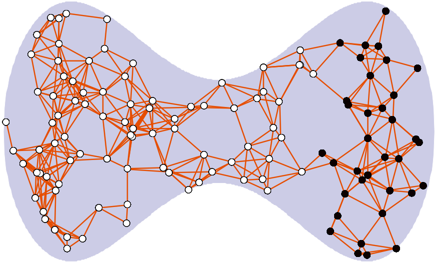
converge as the number of data points $n \to \infty$ to a minimizer of

$$C(u) = \frac{TV(u, \rho^2)}{\min_{c\in\mathbb{R}} \int_D |u(x) - c|\rho(x)dx} \quad ?$$

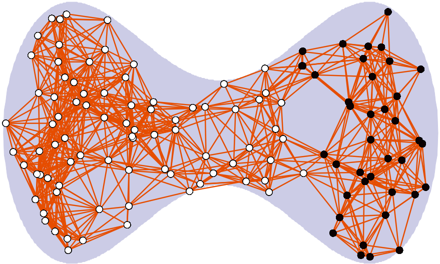
Question 1: For what scaling of $\varepsilon(n)$ can this hold?
Question 2: What is the topology for which $u^n \longrightarrow u$?



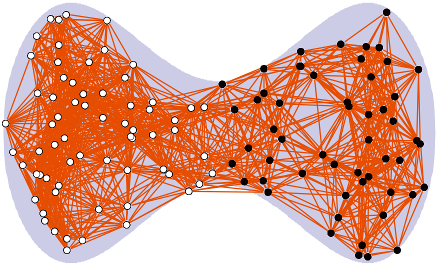
$n = 120, \varepsilon = 0.15$



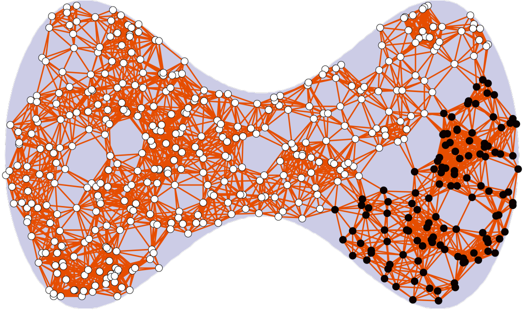
$n = 120, \varepsilon = 0.20$



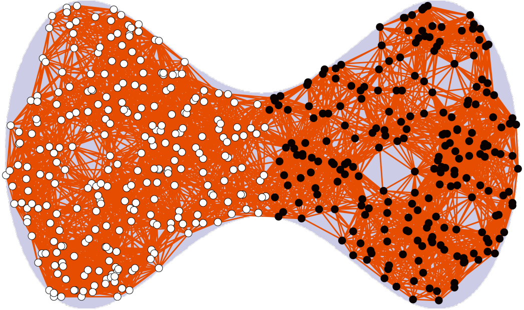
$n = 120, \varepsilon = 0.30$



$n = 120, \varepsilon = 0.40$



$n = 500, \varepsilon = 0.14$



$n = 500, \varepsilon = 0.2$

What was known

Consistency results in statistics/machine learning

- *Arias Castro, Pelletier, and Pudlo 2012* - partial results on the problem
- *Pollard 1981* - k -means
- *Hartigan 1981* - single linkage
- *Belkin and Niyogi 2006* - Laplacian eigenmaps
- *von Luxburg, Belkin, and Bousquet 2004, 2008* - spectral embedding

What was known

Consistency results in statistics/machine learning

- *Arias Castro, Pelletier, and Pudlo 2012* - partial results on the problem
- *Pollard 1981* - k -means
- *Hartigan 1981* - single linkage
- *Belkin and Niyogi 2006* - Laplacian eigenmaps
- *von Luxburg, Belkin, and Bousquet 2004, 2008* - spectral embedding

Calculus of Variations
Discrete to continuum for functionals on grids: *Braides 2010, Braides and Yip 2012, Chambolle, Giacomini and Lussardi 2012, Gobbino and Mora 2001, Van Gennip and Bertozzi 2014*

Γ-Convergence

(Y, d_Y) - metric space, $F_n : Y \to [0, \infty]$

Definition

The sequence $\{F_n\}_{n\in\mathbb{N}}$ Γ-**converges** (w.r.t d_Y) to $F : Y \to [0, \infty]$ if:

Liminf inequality: For every $y \in Y$ and whenever $y_n \to y$

$$\liminf_{n\to\infty} F_n(y_n) \geq F(y),$$

Limsup inequality: For every $y \in Y$ there exists $y_n \to y$ such that

$$\limsup_{n\to\infty} F_n(y_n) \leq F(y).$$

Definition (Compactness property)

$\{F_n\}_{n\in\mathbb{N}}$ satisfies the **compactness property** if

$$\left.\begin{array}{l} \{y_n\}_{n\in\mathbb{N}} \text{ bounded and} \\ \{F_n(y_n)\}_{n\in\mathbb{N}} \text{ bounded} \end{array}\right\} \implies \{y_n\}_{n\in\mathbb{N}} \text{ has convergent subsequence}$$

Proposition: Convergence of minimizers

Γ-convergence and Compactness imply: If y_n is a minimizer of F_n and $\{y_n\}_{n\in N}$ is bounded in Y then along a subsequence

$$y_n \to y \qquad \text{as } n \to \infty$$

and

$$y \text{ is a minimizer of } F.$$

In particular, if F has a unique minimizer, then a sequence $\{y_n\}_{n\in\mathbb{N}}$ converges to the unique minimizer of F.

Consistency of clustering III

Show that

$$GC_{n,\varepsilon_n}(u^n) = \frac{1}{n} \frac{\frac{1}{\varepsilon_n} \sum_{i,j} \eta_{\varepsilon_n}(X_i - X_j) \, |u_i^n - u_j^n|}{\min_{c \in \mathbb{R}} \sum_i |u_i^n - c|}$$

Γ**-converge** as the number of data points $n \to \infty$, and $\varepsilon_n \to 0$ at certain rate to

$$F(u) = \frac{\sigma \, TV(u, \rho^2)}{\min_{c \in \mathbb{R}} \int_D |u(x) - c| \rho(x) dx}$$

and show that compactness property holds.

Questions

1. For what scaling of $\varepsilon(n)$ can this hold?
2. What is the topology for $u^n \longrightarrow u$?

Consistency of graph total variation

Show that

$$GTV_{n,\varepsilon_n}(u^n) = \frac{1}{\varepsilon_n\, n^2} \sum_{i,j} \eta_{\varepsilon_n}(X_i - X_j)\, |u_i^n - u_j^n|$$

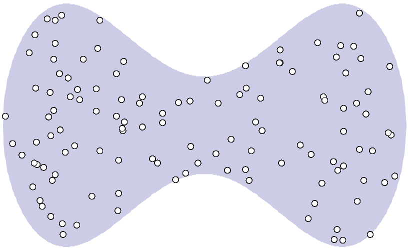
Γ-converge to $\sigma TV(u, \rho^2)$, as the number of data points $n \to \infty$, and $\varepsilon_n \to 0$ at certain rate and show that compactness property holds.

Questions

1. For what scaling of $\varepsilon(n)$ can this hold?
2. What is the topology for $u^n \longrightarrow u$?

Topology

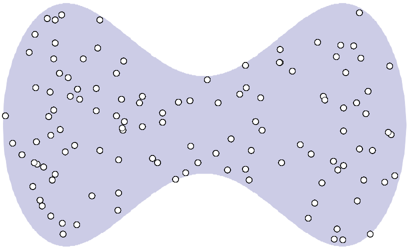
Consider domain D and $V_n = \{X_1, \ldots, X_n\}$ random i.i.d points.



- 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 $\nu_n = \frac{1}{N}\sum_{i=1}^{n} \delta_{X_i}$.

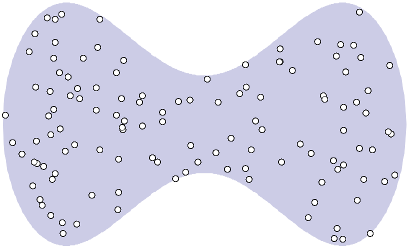
Consider domain D and $V_n = \{X_1, \ldots, X_n\}$ random i.i.d points.



- How to compare $u_n \in L^1(\nu_n)$ and $u \in L^1(D)$ in a way consistent with L^1 topology?

Topology

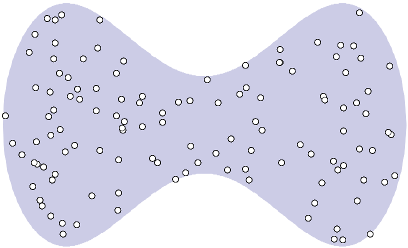
Consider domain D and $V_n = \{X_1, \ldots, X_n\}$ random i.i.d points.



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

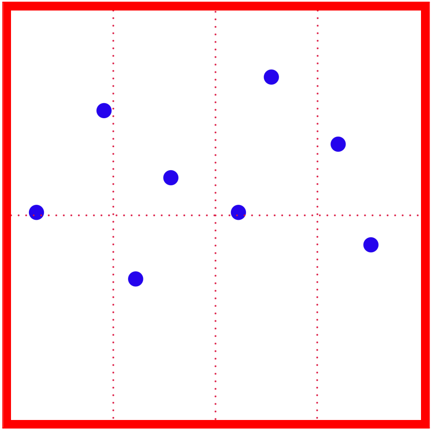
Note that $u \in L^1(\nu)$ and $u_n \in L^1(\nu_n)$, where $\nu_n = \frac{1}{N} \sum_{i=1}^{n} \delta_{X_i}$.

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

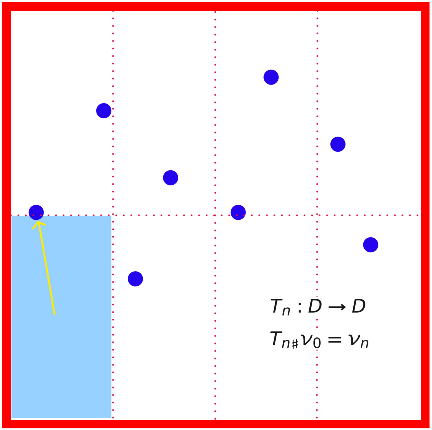


- How to compare $u_n \in L^1(\nu_n)$ and $u \in L^1(D)$ in a way consistent with L^1 topology?

An idea: Divide the domain D into n sets of the same ν measure and to each piece associate a point X_i. That is, consider a map $T_n : D \to D$ such that $T_{\#}\nu = \nu_n$.

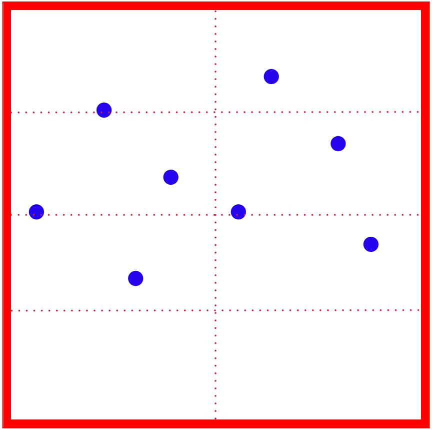


Divide the domain D into n pieces and to each piece associate a point X_i. That is, consider a map $T_n : D \to D$ such that $T_{n\sharp}\nu = \nu_n$.

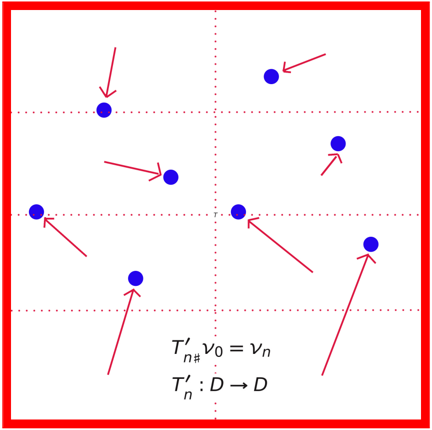


To compare $u \in L^1(\nu)$ and $u_n \in L^1(\nu_n)$ we compare $u_n \circ T_n$ and u in $L^1(\nu)$.

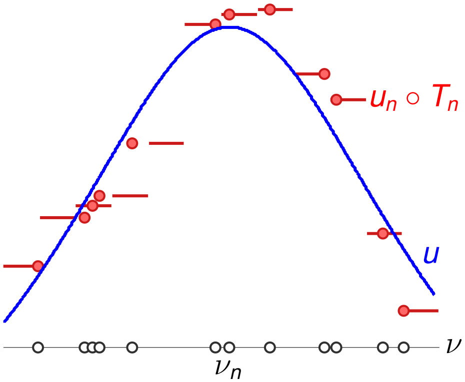
A different partition:



A different partition:



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

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^1 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

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

If $T_\sharp \nu = \sigma$ then $\pi = (I \times T)_\sharp \nu \in \Pi(\nu, \sigma)$ and the integral becomes

$$\int |T(x) - x|^p + |g(T(x)) - f(x)|^p d\nu(x)$$

Lemma

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

TL^1 convergence

- $(\nu, f_n) \xrightarrow{TL^p} (\nu, f)$ iff $f_n \xrightarrow{L^1(\nu)} f$
- $(\nu_n, f_n) \xrightarrow{TL^p} (\nu, f)$ 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

$$\int_{D \times D} |x - y|^p d\pi_n(x, y) \longrightarrow 0 \quad \text{as } n \to \infty. \tag{1}$$

We call a sequence of transportation plans $\pi_n \in \Pi(\nu_n, \nu)$ **stagnating** if it satisfies (1).

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

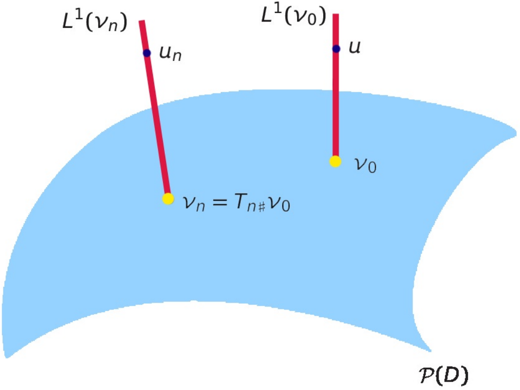
TFAE:

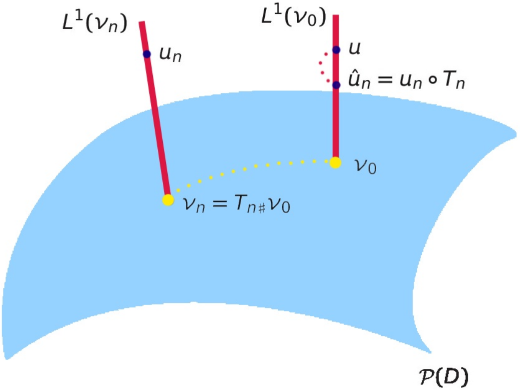
1. $(\nu_n, f_n) \xrightarrow{TL^p} (\nu, f)$ as $n \to \infty$.
2. $\nu_n \rightharpoonup \nu$ and **there exists** a stagnating sequence of transportation plans $\{\pi_n\}_{n\in\mathbb{N}}$ for which

$$\iint_{D\times D} |f(x) - f_n(y)|^p \, d\pi_n(x,y) \to 0, \ as\ n \to \infty. \tag{2}$$

3. $\nu_n \rightharpoonup \nu$ and **for every** stagnating sequence of transportation plans π_n, (2) holds.

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




$L^1(\nu_n)$
u_n
$L^1(\nu_0)$
u
$\hat{u}_n = u_n \circ T_n$
ν_0
$\nu_n = T_{n\sharp}\nu_0$
$\mathcal{P}(D)$

Consistency

$$GTV_{n,\varepsilon_n}(u^n) = \frac{1}{\varepsilon_n\, n^2} \sum_{i,j} \eta_{\varepsilon_n}(X_i - X_j)\, |u_i^n - u_j^n|$$

Γ-convergence of Total Variation (García Trillos and S.)

Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of positive numbers converging to 0 satisfying

$$\lim_{n\to\infty} \frac{(\log n)^{3/4}}{n^{1/2}} \frac{1}{\varepsilon_n} = 0 \;\; \text{if } d = 2,$$
$$\lim_{n\to\infty} \frac{(\log n)^{1/d}}{n^{1/d}} \frac{1}{\varepsilon_n} = 0 \;\; \text{if } d \geq 3.$$

Then, GTV_{n,ε_n} Γ-converge to $\sigma\, TV(\,\cdot\,, \rho^2)$ as $n \to \infty$ in the TL^1 sense, where σ depends explicitly on η.

Consistency

Γ-convergence of Perimeter

The conclusions hold when all of the functionals are restricted to characteristic functions of sets. That is, the graph perimeters Γ-converge to the continuum perimeter.

Compactness

With the same conditions on ε_n as before, if

$$\sup_{n\in\mathbb{N}} \|u_n\|_{L^1(D,\nu_n)} < \infty,$$

and

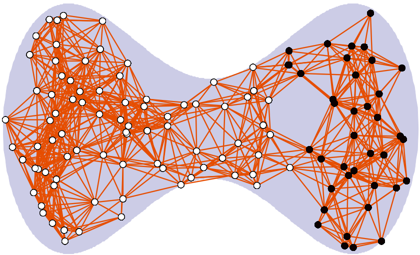
$$\sup_{n\in\mathbb{N}} GTV_{n,\varepsilon_n}(u_n) < \infty,$$

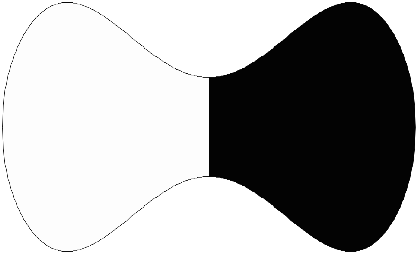
then $\{u_n\}_{n\in N}$ is TL^1-precompact.

Recall:

$$GC_{n,\varepsilon_n}(u^n) = \frac{1}{n} \frac{\frac{1}{\varepsilon_n} \sum_{i,j} \eta_{\varepsilon_n}(X_i - X_j) \, |u_i^n - u_j^n|}{\min_{c\in\mathbb{R}} \sum_i |u_i^n - c|}$$

$$C(u) = \frac{\sigma TV(u, \rho^2)}{\min_{c\in\mathbb{R}} \int_D |u(x) - c| \rho(x) dx}$$





Consistency of Cheeger Cuts

Recall:

$$GC_{n,\varepsilon_n}(u^n) = \frac{1}{n} \frac{\frac{1}{\varepsilon_n} \sum_{i,j} \eta_{\varepsilon_n}(X_i - X_j) \, |u_i^n - u_j^n|}{\min_{c \in \mathbb{R}} \sum_i |u_i^n - c|}$$

$$C(u) = \frac{\sigma TV(u, \rho^2)}{\min_{c \in \mathbb{R}} \int_D |u(x) - c| \rho(x) dx}$$

Consistency of Cheeger Cuts (von Brecht, García Trillos, Laurent, S.)

For the same conditions on ε_n as before, with probability one:

$$GC_{n,\varepsilon_n} \xrightarrow{\Gamma} C \qquad \text{w.r.t. } TL^1 \text{ metric.}$$

Moreover, for any sequence of sets $E_n \subseteq \{X_1, \ldots, X_n\}$ of almost minimizers of the Cheeger energy, every subsequence has a convergent subsequence (in the TL^1 sense) to a minimizer of the Cheeger energy on the domain D.

Comment of ε_n

- We require

$$\lim_{n\to\infty} \frac{(\log n)^{3/4}}{n^{1/2}} \frac{1}{\varepsilon_n} = 0 \;\; \textit{if } d = 2,$$
$$\lim_{n\to\infty} \frac{(\log n)^{1/d}}{n^{1/d}} \frac{1}{\varepsilon_n} = 0 \;\; \textit{if } d \geq 3.$$

- 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

$$\lim_{n\to\infty} \frac{(\log n)^{3/4}}{n^{1/2}} \frac{1}{\varepsilon_n} = 0 \;\; \textit{if } d = 2,$$

$$\lim_{n\to\infty} \frac{(\log n)^{1/d}}{n^{1/d}} \frac{1}{\varepsilon_n} = 0 \;\; \textit{if } d \geq 3.$$

- 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. *Penrose (1999); Gupta and Kumar (1999); Goel,Rai and Krishnamachari (2004).*
  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.

Hint about the proof

Assume that $u_n \xrightarrow{TL^1} u$ as $n \to \infty$.
There exists $T_{n\sharp}\nu = \nu_n$ stagnating ($\int |x - T_n(x)| d\nu(x) \to 0$).

$$GTV_{n,\varepsilon_n}(u^n) = \frac{1}{\varepsilon_n} \int_{D\times D} \eta_{\varepsilon_n}(\tilde{x} - \tilde{y}))\, |u_n(\tilde{x}) - u_n(\tilde{y})|\, d\nu_n(\tilde{x}) d\nu_n(\tilde{y})$$
$$= \frac{1}{\varepsilon_n} \int_{D\times D} \eta_{\varepsilon_n}(T_n(x) - T_n(y))\, |u_n \circ T_n(x) - u_n \circ T_n(y)|\, \rho(x)\rho(y) dx dy$$

Define $TV_\varepsilon(u;\rho) := \frac{1}{\varepsilon} \int_{D\times D} \eta_\varepsilon(x-y)|u(x)-u(y)|\rho(x)\rho(y)dxdy$.

- $TV_\varepsilon \xrightarrow{\Gamma} TV(\,\cdot\,, \rho^2)$ wrt $L^1(\nu_0)$ metric.
  (*Alberti-Bellettini, Chambolle-Giacomini-Lussardi, Savin-Valdinocci, Ponce*)
- If $|T_n(x) - x| \ll \varepsilon_n$ then one may be able to compare $GTV_{n,\varepsilon_n}(u^n)$ and $TV_\varepsilon(u_n \circ T_n; \rho)$.

$\infty-$transportation distance:

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

- There exists a minimizer $\pi \in \Pi(\mu,\nu)$.
- 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

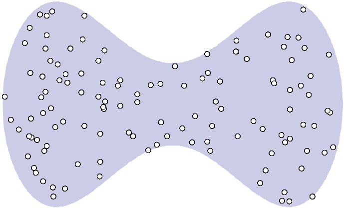
$$d_\infty(\mu,\nu) = \min_{\sigma-\text{permutation}} \max_i |x_i - y_{\sigma(i)}|.$$

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

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

∞-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)*

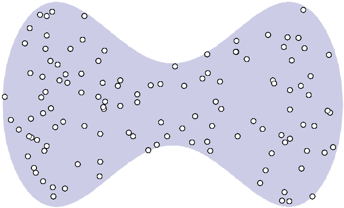


Theorem

There are constants $c > 0$ and $C > 0$ (only 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:

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

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:

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

- $V_n = \{X_1, \ldots, X_n\}$, similarity matrix W, as before:
$$W_{ij} := \frac{1}{\varepsilon^d}\,\eta\left(\frac{|X_i - X_j|}{\varepsilon}\right).$$
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 $L = D - W$, where $D = \mathrm{diag}(d_1, \ldots, d_n)$.
- To partition the point cloud into two clusters, consider the eigenvector corresponding to 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 of Spectral Clustering

von Luxburg, Belkin, Bousquet '08, Belkin-Nyogi '07, Ting, Huang, Jordan '10, Singer, Wu '13, Burago, Ivanov, Kurylev '14, Shi, Sun '15

$$u_n := \arg\min \left\{ \sum_{i,j} W_{ij} |u(X_i) - u(X_j)|^2 \;:\; \sum_i u(X_i) = 0, \; \|u\|_2 = 1 \right\}$$

- Suppose $X_1, \dots, X_n, \dots$ are i.i.d samples of a distribution with density ρ. Then, for $\varepsilon_n \to 0$ as before

$$u_n \xrightarrow{TL^1} u$$

where u is eigenfunction, corresponding to second eigenvalue, of

$$-\frac{\operatorname{div}(\rho^2 \nabla u)}{\rho} = \lambda_2 u \quad \text{in } D$$

$$\frac{\partial u}{\partial n} = 0 \qquad \text{on } \partial D.$$

Consistency of Spectral Clustering II

$$u_n^k = \arg\min \left\{ \sum_{i,j} W_{ij} |u(X_i) - u(X_j)|^2 \, : \, \sum_i u(X_i) u_n^m(X_i) = 0 \; (\forall m < k), \|u\|_2 = 1 \right\}$$

- Suppose $X_1, \ldots, X_n, \ldots$ are i.i.d samples of a distribution with density ρ. Then, for $\varepsilon_n \to 0$ as before

$$u_n^k \xrightarrow{TL^1} u^k$$

where u^k is eigenfunction, corresponding to k-th eigenvalue, of

$$-\frac{1}{\rho}\,\mathrm{div}(\rho^2 \nabla u^k) = \lambda_k u^k \quad \text{in } D$$

$$\frac{\partial u^k}{\partial n} = 0 \qquad \text{on } \partial D.$$

Consistency of spectral clustering

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 (nonzero) eigenvalues of L.
- Define the matrix $U \in \mathbb{R}^{k \times n}$, where the i-th row of U is the vector u_i.
- For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the i-th column of U.
- 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 $G_1, \ldots, G_k$.

Theorem. Let $G_1^n, \ldots G_k^n$ be the clusters above. Let $\nu_i^n = \nu_{n \llcorner G_i^n}$ (the restriction of empirical measure to clusters) for $i = 1, \ldots, k$. Then $(\nu_1^n, \ldots, \nu_k^n)$ is precompact with respect to weak convergence of measures and converges along a subsequence to $(\nu_1, \ldots, \nu_k) = (\nu_{\llcorner G_1}, \ldots, \nu_{\llcorner G_k})$ where $G_1, \ldots, G_k$ is a continuum spectral clustering of ν.

Normalized Graph Laplacian

- As before: $W_{ij} := \frac{1}{\varepsilon^d}\eta\left(\frac{|X_i - X_j|}{\varepsilon}\right)$, $d_i = \sum_j W_{i,j} = \sum_j \eta_\varepsilon(|X_i - X - j|)$.
- Dirichlet energy of $u_n : V_n \to \mathbb{R}$ is

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

- Associated operator is the normalized graph laplacian $D^{-1/2}LD^{-1/2} = I - D^{-1/2}WD^{-1/2}$, where $D = \mathrm{diag}(d_1, \ldots, d_n)$.
- To partition the point cloud into two clusters, consider the eigenvector corresponding to second eigenvalue:

$$u_n := \arg\min\left\{\sum_{i,j} W_{ij}\left|\frac{u_n(X_i)}{\sqrt{d_i}} - \frac{u_n(X_j)}{\sqrt{d_j}}\right|^2 : \sum_i u(X_i) = 0,\ \|u\|_2 = 1\right\}$$

$$u_n^k = \arg\min \left\{ \sum_{i,j} \left| \frac{u_n(X_i)}{\sqrt{d_i}} - \frac{u_n(X_j)}{\sqrt{d_j}} \right|^2 \; : \; \sum_i u(X_i) u_n^m(X_i) = 0 \; (\forall m < k), \|u\|_2 = 1 \right\}$$

- Suppose $X_1, \dots, X_n, \dots$ are i.i.d samples of a distribution with density ρ. Then, for $\varepsilon_n \to 0$ as before

$$u_n^k \xrightarrow{TL^1} u^k$$

where u^k is eigenfunction, corresponding to k-th eigenvalue, of

$$-\frac{1}{\rho^{3/2}} \nabla \cdot \left(\rho^2 \nabla \left(\frac{u^k}{\sqrt{\rho}} \right) \right) = \lambda_k u^k \quad \text{in } D$$
$$\frac{\partial (u^k/\sqrt{\rho})}{\partial n} = 0 \qquad \text{on } \partial D.$$

Future directions and ongoing work

- Jointly with Hein, Garciá Trillos, and Gerlach
  - Consistency of graph cuts and spectral clustering on manifolds
  - Error estimates for consistency of convex functionals (like the Dirichlet functional)
- Mumford–Shah functional on graphs
- Finding better ways to approximate the functionals

L^∞ to d_∞

Theorem

Let $D \subset \mathbb{R}^d$ be a bounded, connected, open set with Lipschitz boundary. Let ν_1, ν_2 be measures on D of the same total mass: $\nu_1(D) = \nu_2(D)$. Assume the measures are absolutely continuous with respect to Lebesgue measure and let ρ_1 and ρ_2 be their densities. Furthermore assume that for some $\lambda > 1$, for all $x \in D$

$$\frac{1}{\lambda} \leq \rho_i(x) \leq \lambda \quad \text{for } i = 1, 2. \tag{4}$$

Then, there exists a constant $C(\lambda, D)$ depending only on λ and D such that for all ν_1, ν_2 as above

$$d_\infty(\nu_1, \nu_2) \leq C(\lambda, D) \|\rho_1 - \rho_2\|_{L^\infty(D)}.$$