From clustering with graph cuts to isoperimetric inequalities: quantitative convergence rates of Cheeger cuts on data clouds

Nicolás García Trillos
University of Wisconsin-Madison
Joint work with Ryan Murray (NCSU) and Matthew Thorpe (U Manchester)

IPAM High Dimensional Hamilton-Jacobi PDEs Workshop 2
April 24th 2020

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

- Unsupervised learning: Find coarse structure of X (find meaningful clusters).

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

- Unsupervised learning: Find coarse structure of X (find meaningful clusters).

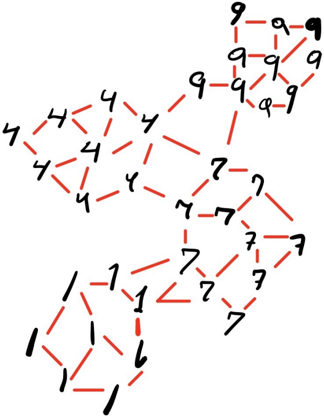
Graph based learning

Goal: Given a data set $X = \{x_1, \ldots, x_n\}$ and similarity matrix $\{\omega_{ij}\}_{ij}$:

- Find coarse structure of X (find meaningful clusters).

Graph based learning

Given $G = (X, \omega)$:

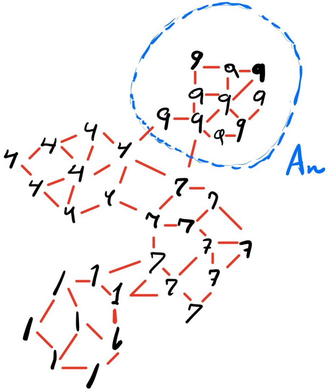


What can we do with a graph?

Define **graph cut** (**graph perimeter**) for arbitrary $A_n \subseteq X$:

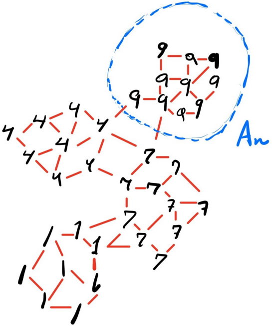
$$Cut(A_n) := \sum_{x_i \in A_n} \sum_{x_j \notin A_n} \omega_{ij} = \sum_{i,j} \omega_{ij} |\mathbb{1}_{A_n}(x_i) - \mathbb{1}_{A_n}(x_j)|$$

What can we do with a graph?



$$Cut(A_n) := \sum_{x_i \in A_n} \sum_{x_j \notin A_n} \omega_{ij} = \sum_{i,j} \omega_{ij} |\mathbb{1}_{A_n}(x_i) - \mathbb{1}_{A_n}(x_j)|$$

What can we do with a graph?



$$Cut(A_n) := \sum_{x_i \in A_n} \sum_{x_j \notin A_n} \omega_{ij} = \sum_{i,j} \omega_{ij} |\mathbb{1}_{A_n}(x_i) - \mathbb{1}_{A_n}(x_j)|$$

Fundamental object that appears in Computer Science and Math (e.g. min cut max flow theorem), but also used for clustering...

Two-way clustering

The **Cheeger** cut problem (an example of **balanced graph cut**):

$$\min_{A_n \subseteq \mathcal{M}_n} \frac{Cut(A_n)}{\min\{\mu_n(A_n), \mu_n(A_n^c)\}}$$

Two-way clustering

The **Cheeger** cut problem (an example of **balanced graph cut**):

$$\min_{A_n \subseteq \mathcal{M}_n} \frac{Cut(A_n)}{\min\{\mu_n(A_n), \mu_n(A_n^c)\}}$$

Computational complexity: **NP hard**.

Relaxations

First notice that,

$$\sum_{i,j} \omega_{ij} |\mathbb{1}_{A_n}(x_i) - \mathbb{1}_{A_n}(x_j)| = \sum_{i,j} \omega_{ij} |\mathbb{1}_{A_n}(x_i) - \mathbb{1}_{A_n}(x_j)|^2$$

Second, consider problem over all functions $f : X \to \mathbb{R}$:

$$\min_f \frac{\sum_{i,j} \omega_{ij} |f(x_i) - f(x_j)|^2}{\sum_{i=1}^n (f(x_i) - \frac{1}{n} \sum_{j=1}^n f(x_j))^2}$$

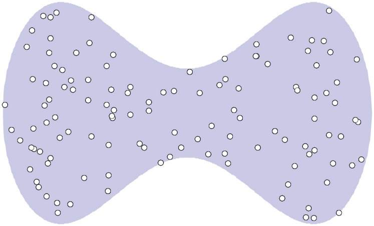
Eigenvalue problem for **graph Laplacian** Δ_n:

$$\Delta_n f(x_i) := \sum_{ij} \omega_{ij} (f(x_i) - f(x_j)).$$

Spectral clustering: Ng et al (2002), von Luxburg (2007).

What if the data was already embedded in Euclidean space?

$X := \{x_1, \ldots, x_n\} \subseteq \mathcal{M} \subseteq \mathbb{R}^d$ with $m \ll d$ (**the manifold assumption**).



Here

$$\omega_{ij} = \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right), \quad \text{e.g. } \eta(t) := \begin{cases} 1 & \text{if } t \leq 1 \\ 0 & \text{else} \end{cases}$$

Here

$$\omega_{ij} = \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right), \quad \text{e.g. } \eta(t) := \begin{cases} 1 & \text{if } t \leq 1 \\ 0 & \text{else} \end{cases}$$

Here

$$\omega_{ij} = \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right), \quad \text{e.g. } \eta(t) := \begin{cases} 1 & \text{if } t \leq 1 \\ 0 & \text{else} \end{cases}$$

Other families of proximity graphs: *k*-NN graphs, self-tuning graphs, graphs based on polar curvature of points (e.g. Chen and Lerman 2007).

Manifold Learning

What is the behavior of algorithms as $n \to \infty$ (and $\varepsilon \to 0$)?

$$\omega_{ij} = \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right), \quad \text{e.g. } \eta(t) := \begin{cases} 1 & \text{if } t \leq 1 \\ 0 & \text{else} \end{cases}$$

Manifold Learning

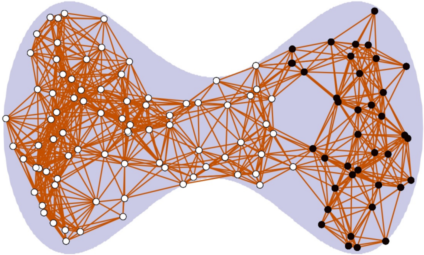
What is the behavior of algorithms as $n \to \infty$ (and $\varepsilon \to 0$)?

$$\omega_{ij} = \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right), \qquad \text{e.g. } \eta(t) := \begin{cases} 1 & \text{if } t \leq 1 \\ 0 & \text{else} \end{cases}$$

Manifold Learning for **graph Laplacian**:

- Towards a theoretical foundation for Laplacian based methods. Belkin and Niyogi (2005).
- Diffusion maps. Coifman et al (2005).
- From graph to manifold Laplacian: the convergence rate. Singer (2006).
- Consistency of Spectral clustering. von Luxburg, Belkin, and Bousquet (2007).
- . . .
- A graph-discretization of the Laplace-Beltrami operator. Burago et al (2014).
- A variational approach to the consistency of spectral clustering. NGT and Slepčev (2015).
- Error estimates for spectral convergence of the graph Laplacian on random geometric graphs towards the Laplace-Beltrami operator. NGT et al (2018).
- Improved spectral convergence rates for graph Laplacians on epsilon and k-NN graphs. Calder and NGT (2019).

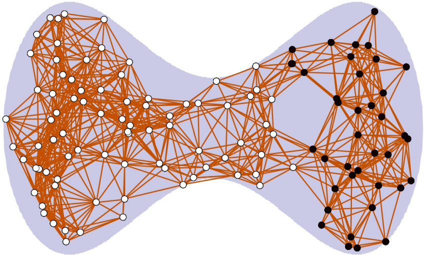
But let's go back to the un-relaxed Cheeger cut problem...



$$A_n^* = \operatorname{argmin}_{A_n \subseteq X} \frac{GTV_{n,\varepsilon}(\mathbb{1}_{A_n})}{\min\{\mu_n(A_n), \mu_n(A_n^c)\}}$$

$$\begin{aligned} GTV_{n,\varepsilon}(\mathbb{1}_{A_n}) &:= \frac{1}{n^2 \varepsilon^{m+1}} \sum_{i,j} \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right) |\mathbb{1}_{A_n}(x_i) - \mathbb{1}_{A_n}(x_j)| \\ &= \frac{1}{n^2 \varepsilon^{m+1}} Cut(A_n) \end{aligned}$$

Consistency question

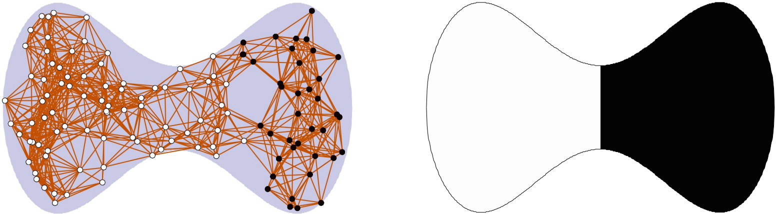


$$A_n^* = \mathrm{argmin}_{A_n \subseteq X} \frac{GTV_{n,\varepsilon}(\mathbb{1}_{A_n})}{\min\{\mu_n(A_n), \mu_n(A_n^c)\}}$$

$X = \{x_1, \ldots, x_n\} \sim_{i.i.d.} \mu$ where μ is a distribution over manifold $\mathcal{M}$. What happens as $n \to \infty$ and simultaneously $\varepsilon \to 0$?

Consistency question

$X = \{x_1, \dots, x_n\} \sim_{i.i.d.} \mu$ where μ is a distribution over manifold $\mathcal{M}$.



What happens as $n \to \infty$ and simultaneously $\varepsilon \to 0$?

What are the difficulties?

$$\min_{A_n \subseteq X} \frac{GTV_{n,\varepsilon}(\mathbb{1}_{A_n})}{\min\{\mu_n(A_n), \mu_n(A_n^c)\}}$$

1. Non-linear, non-convex, random variational problem.
2. Optimizing over **all** subsets A_n of X.

What are the difficulties?

$$\min_{A_n \subseteq X} \frac{GTV_{n,\varepsilon}(\mathbb{1}_{A_n})}{\min\{\mu_n(A_n), \mu_n(A_n^c)\}}$$

1. Non-linear, non-convex, random variational problem.
2. Optimizing over **all** subsets A_n of X.
   Arias-Castro et al: consider A_ns that are restrictions of "nice" subsets of $\mathcal{M}$; for this family one can bound VC dimension.

The classical approach from statistical learning

Try to estimate:

$$\sup_{f \in \mathcal{F}} |\mathcal{E}_n(f) - \mathcal{E}(f)|$$

using **capacity measure** of class of functions $\mathcal{F}$.

The classical approach from statistical learning

Try to estimate:

$$\sup_{f\in\mathcal{F}}|\mathcal{E}_n(f)-\mathcal{E}(f)|$$

using **capacity measure** of class of functions $\mathcal{F}$. In Arias-Castro:

$$\mathcal{F}=\{\mathbb{1}_A\ :\ A \text{ is "nice" enough subset of } \mathcal{M}\}$$

$$\mathcal{E}_n(f):=\frac{GTV_{n,\varepsilon}(\mathbb{1}_{A_n})}{\min\{\mu_n(A_n),\mu_n(A_n^c)\}},\quad A_n:=A\cap X$$

$$\mathcal{E}(f):=\frac{TV_\varepsilon(\mathbb{1}_A)}{\min\{\mu(A),\mu(A^c)\}}.$$

Where $TV_\varepsilon(f)$ is the **non-local** TV seminorm:

$$TV_\varepsilon(f):=\frac{1}{\varepsilon^{m+1}}\int_{\mathcal{M}}\int_{\mathcal{M}}\eta\left(\frac{|x_i-x_j|}{\varepsilon}\right)|f(x)-f(y)|d\mu(y)d\mu(x)$$

But if we want to analyze the true Cheeger cut problem, this approach can not work.

Need to: “Pass the probabilistic estimates to where they do not harm".

- N. Garcia Trillos, D. Slepcev, J. Von Brecht, T. Laurent, and X. Bresson. *Consistency of cheeger and ratio graph cuts*. JMLR, 17(1):6268-6313, 2016.
- N. Garcia Trillos and D. Slepcev. *Continuum limit of total variation on point clouds*. ARMA, 220(1):193-241, 2016.

Asymptotic consistency

Theorem: Suppose that ε scales like:

$$\frac{\log(n)^{p_m}}{n^{1/m}} \ll \varepsilon \ll 1, \quad p_m = \begin{cases} 3/4 \text{ if } m = 2 \\ 1/m \text{ if } m \geq 3 \end{cases}$$

Then, with probability one,

$$\lim_{n\to\infty} \mathcal{C}_{n,\varepsilon} = \sigma_\eta \mathcal{C}, \qquad \text{and } A_n^* \to_{TL^1} A^*.$$

Where $\mathcal{C}$ and A^* solve:

$$\min_{A\subseteq\mathcal{M}} \frac{\mathrm{TV}(\mathbb{1}_A)}{\min\{\mu(A), \mu(A^c)\}}$$

and

$$\sigma_\eta := \int_{\mathbb{R}^m} \eta(|z|)|z_1| dz,$$

$$TV(f) = \sup \left\{ \int_{\mathcal{M}} f \mathrm{div}(V) dx \ : \ V \in C_c^{\infty}(\mathcal{M}), \quad |V(x)| \leq 1, \quad \forall x \right\}.$$

$$\mathrm{Per}(A) := TV(\mathbb{1}_A)$$

These expressions reduce to

$$TV(f) = \int_{\mathcal{M}} |\nabla f(x)| dx,$$

for f **smooth**, and

$$\mathrm{Per}(A) = \int_{\partial A} d\mathcal{H}^{m-1}(x).$$

for A with **smooth boundary**.

Asymptotic consistency

Theorem: Suppose that ε scales like:

$$\frac{\log(n)^{p_m}}{n^{1/m}} \ll \varepsilon \ll 1, \quad p_m = \begin{cases} 3/4 \text{ if } m = 2 \\ 1/m \text{ if } m \geq 3 \end{cases}$$

Then, with probability one,

$$\lim_{n\to\infty} \mathcal{C}_{n,\varepsilon} = \sigma_\eta \mathcal{C}, \qquad \text{and } A_n^* \to_{TL^1} A^*.$$

Where $\mathcal{C}$ and A^* solve:

$$\min_{A\subseteq\mathcal{M}} \frac{\mathrm{TV}(\mathbb{1}_A)}{\min\{\mu(A), \mu(A^c)\}}$$

Remarks:

1. Proof: relied on Γ-convergence... and **NO rates** of convergence.
2. Penrose and Muller (2018) improved the result for $m = 2$.

Convergence rates for Cheeger constants $\mathcal{C}_{n,\varepsilon}$

Upper bound: $\mathcal{C}_{n,\varepsilon} \lesssim \sigma_\eta \mathcal{C}$

Fix $f := \mathbb{1}_{A^*}$. Concentration inequalities to estimate:

$$\mathbb{P}\left(|GTV_{n,\varepsilon}(f) - \mathbb{E}(GTV_{n,\varepsilon}(f))| \geq t\right).$$

Also,

$$\mathbb{E}(GTV_{n,\varepsilon}(f)) = TV_\varepsilon(f)$$

and

$$\sigma_\eta TV(\mathbb{1}_{A^*}) \gtrsim TV_\varepsilon(\mathbb{1}_{A^*}) = GTV_{n,\varepsilon}(\mathbb{1}_{A^*}) + (TV_\varepsilon(\mathbb{1}_{A^*}) - GTV_{n,\varepsilon}(\mathbb{1}_{A^*}))\,.$$

Upper bound: $\mathcal{C}_{n,\varepsilon} \lesssim \sigma_\eta \mathcal{C}$

Combine:

$$GTV_{n,\varepsilon}(\mathbb{1}_{A^*}) \leq \sigma_\eta TV(\mathbb{1}_{A^*}) + error_{n,\varepsilon}$$

with

$$|\mu_n(A^*) - \mu(A^*)| \leq error_{n,\varepsilon}$$

to conclude:

$$\begin{aligned} \mathcal{C}_{n,\varepsilon} \leq \frac{GTV_{n,\varepsilon}(\mathbb{1}_{A^*})}{\min\{\mu_n(A^*), \mu_n(A^{*c})\}} &\leq \frac{\sigma_\eta TV(\mathbb{1}_{A^*})}{\min\{\mu(A^*), \mu(A^{*c})\}} + error_{n,\varepsilon} \\ &= \sigma_\eta \mathcal{C} + error_{n,\varepsilon} \end{aligned}$$

Lower bound: $\sigma_\eta \mathcal{C} \lesssim \mathcal{C}_{n,\varepsilon}$

Proposition:[NGT, Murray, Thorpe 20'] As long as $\left(\frac{log(n)}{n}\right)^{1/m} \ll \varepsilon \ll 1$, w.v.h.p. **for all** $A_n \subset X$ there is $A \subset \mathcal{M}$ such that:

1. $\mu_n(A_n \triangle A) \leq GTV_{n,\varepsilon}(\mathbb{1}_{A_n}) \cdot error_{n,\varepsilon}$.
2. $\sigma_\eta TV(\mathbb{1}_A) \leq (1 + error_{n,\varepsilon}) GTV_{n,\varepsilon}(\mathbb{1}_{A_n})$.

Pass the probabilistic estimates to where they do not harm... (typically involves a "good" interpolation map)

Lower bound: $\sigma_\eta \mathcal{C} \lesssim \mathcal{C}_{n,\varepsilon}$

Proposition:[NGT, Murray, Thorpe 20'] As long as $\left(\frac{log(n)}{n}\right)^{1/m} \ll \varepsilon \ll 1$, w.v.h.p., **for all** $A_n \subset X$ there is $A \subset \mathcal{M}$ such that:

1. $\mu_n(A_n \triangle A) \leq GTV_{n,\varepsilon}(\mathbb{1}_{A_n}) \cdot error_{n,\varepsilon}$.
2. $TV(\sigma_\eta \mathbb{1}_A) \leq (1 + error_{n,\varepsilon}) GTV_{n,\varepsilon}(\mathbb{1}_{A_n})$.

Take $A_n := A_n^*$ and proceed as in the upper bound to get:

$$\sigma_\eta \mathcal{C} \leq \mathcal{C}_{n,\varepsilon} + error_{n,\varepsilon}.$$

A collection of (deterministic) inequalities

1. For all $f \in L^1(\mathcal{M})$

$$TV_\varepsilon(f) \leq (1 + C\varepsilon^2)\sigma_\eta TV(f),$$

A collection of inequalities

1. For all $f \in L^1(\mathcal{M})$

$$TV_\varepsilon(f) \leq (1 + C\varepsilon^2)\sigma_\eta TV(f),$$

2. If $f \in C^2(\mathcal{M})$ then,

$$\sigma_\eta TV(f) \leq (1 + C\varepsilon^2) TV_\varepsilon(f) + C\|f\|_{C^2(\mathcal{M})}\varepsilon.$$

A collection of inequalities

1. For all $f \in L^1(\mathcal{M})$
$$TV_\varepsilon(f) \leq (1 + C\varepsilon^2)\sigma_\eta TV(f),$$
2. If $f \in C^2(\mathcal{M})$ then,
$$\sigma_\eta TV(f) \leq (1 + C\varepsilon^2) TV_\varepsilon(f) + C\|f\|_{C^2(\mathcal{M})}\varepsilon.$$
3. (Monotonicity) For all $0 < h \leq a$ small enough, for all $f \in L^1(\mathcal{M})$
$$TV_a(f) \leq C TV_h(f).$$

A collection of inequalities

1. For all $f \in L^1(\mathcal{M})$
$$TV_\varepsilon(f) \leq (1 + C\varepsilon^2)\sigma_\eta TV(f),$$
2. If $f \in C^2(\mathcal{M})$ then,
$$\sigma_\eta TV(f) \leq (1 + C\varepsilon^2) TV_\varepsilon(f) + C\|f\|_{C^2(\mathcal{M})}\varepsilon.$$
3. (Monotonicity) For all $0 < h \leq a$ small enough, for all $f \in L^1(\mathcal{M})$
$$TV_a(f) \leq C TV_h(f).$$
4. For all $0 < h \leq a$ small enough, for all $f \in L^1(\mathcal{M})$
$$\|f - \Lambda_a f\|_{L^1(\mathcal{M})} \leq Ca TV_h(f).$$

A collection of inequalities

1. For all $f \in L^1(\mathcal{M})$
$$TV_\varepsilon(f) \leq (1 + C\varepsilon^2)\sigma_\eta TV(f),$$
2. If $f \in C^2(\mathcal{M})$ then,
$$\sigma_\eta TV(f) \leq (1 + C\varepsilon^2) TV_\varepsilon(f) + C\|f\|_{C^2(\mathcal{M})}\varepsilon.$$
3. (Monotonicity) For all $0 < h \leq a$ small enough, for all $f \in L^1(\mathcal{M})$
$$TV_a(f) \leq CTV_h(f).$$
4. For every $a > 0$ small enough and every $0 < h \leq a$
5. For all $0 < h \leq a$ small enough, for all $f \in L^1(\mathcal{M})$
$$\|f - \Lambda_a f\|_{L^1(\mathcal{M})} \leq CaTV_h(f).$$
6. (Convolution decreases energy) For all $0 < h \leq a$ small enough, for all $f \in L^1(\mathcal{M})$
$$TV_h(\Lambda_a f) \leq (1 + Ca) TV_h(f).$$

Bottom line:
For all $f \in L^1(\mathcal{M}) \cap L^\infty(\mathcal{M})$:

1. $\|\Lambda_a f - f\| \leq a TV_\varepsilon(f)$
2. $\sigma_\eta TV(\Lambda_a f) \leq TV_\varepsilon(f)$

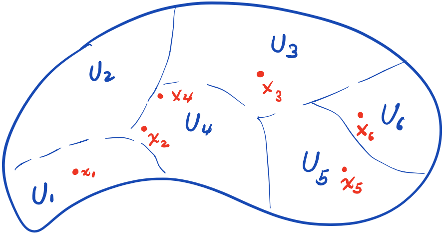
To complete the proof of the proposition we only need to relate $GTV_{n,\varepsilon}(\mathbb{1}_{A_n^*})$ with $TV_\varepsilon(f)$ for some conveniently chosen f (here is where the probabilistic estimate enters).

$$f := \mathbb{1}_{A_n} \circ T_n$$

Here: $T_n : \mathcal{M} \to X$ satisfies $T_{\sharp}\mu = \mu_n$ and is such that

$$\|T_n - Id\|_\infty$$

is as small as possible:



$$U_i = T_n^{-1}(\{x_i\}), \text{ and } \mu(U_i) = 1/n$$

For any such a map we get deterministically:

$$TV_\varepsilon(\mathbb{1}_{A_n} \circ T_n) \lesssim (1 + C\frac{\| T_n - Id\|_\infty}{\varepsilon}) GTV_{n,\varepsilon}(\mathbb{1}_{A_n}).$$

The probabilistic estimate is used to quantify how good T_n can be...

Convergence of Cheeger constants

Theorem:(NGT, Murray, Thorpe 20') As long as $\left(\frac{log(n)}{n}\right)^{1/m} \ll \varepsilon \ll 1$, w.v.h.p.,

$$|\mathcal{C}_{n,\varepsilon} - \sigma_\eta \mathcal{C}| \leq C\left(\frac{\log(n)^{1/m}}{n^{1/m}\varepsilon} + \sqrt[3]{\varepsilon}\right).$$

What about convergence of **minimizers**?

What about convergence of **minimizers**?

... Finding the right notion of convexity

What about convergence of **minimizers**?

... Finding the right notion of convexity: a detour into isoperimetric inequalities.

A classical problem

Q: (in dimension 2): Find a closed plane curve of a given perimeter that encloses the greatest area.

A classical problem

Q: (in dimension 2): Find a closed plane curve of a given perimeter that encloses the greatest area. **A:** The circle!

A classical problem with rich history

Q: (in dimension 2): Find a closed plane curve of a given perimeter that encloses the greatest area. **A:** The circle!

- The greeks.
- J. Steiner (1830's)
- St. Venant (1850's)
- Rayleigh (1870's)
- Poincare (1900's)
- Faber-Krahn (1920's)
- Cartan-Hadamard conjecture.
- Szego-Polya (1930's and 1940's)
- $\vdots$

Isoperimetric function:

$$\mathbb{I}(v) := \inf_{E} \{ \mathrm{Per}(E) : |E| = v \}.$$

Isoperimetric inequality in $\mathbb{R}^m$:

$$C_m |E|^{(m-1)/m} \leq \mathrm{Per}(E).$$

Inequality is equality if and only if E is a ball.

Q: Quantitative version? i.e. If $|E| = v$ and $\mathrm{Per}(E) - \mathbb{I}(v) < \delta$, can we say how close is E from a minimizer?

Q: Quantitative version? i.e. If $|E| = v$ and $\mathrm{Per}(E) - \mathbb{I}(v) < \delta$, can we say how close is E from a minimizer?

- Bernstein (1905)
- Bonesen (1924)
- Fuglede (1989)
- Hall (1992)
- $\vdots$

Sharp quantitative isoperimetric inequality in $\mathbb{R}^m$

Theorem (Fusco, Maggi, Pratelli 2008)

There exist constants $C, \delta > 0$ *such that for any* $|E| = |B^*|$ *and* $|E \ominus B^*| < \delta$ *one has*

$$\mathrm{Per}(E) - \mathrm{Per}(B^*) \geq C\alpha(E)^2.$$

Here, α *is the* **Fraenkel asymmetry***:*

$$\alpha(E) = \inf_{B^*} \{ |E \ominus B^*| : |B^*| = |E|, \mathrm{Per}(B^*) = \mathbb{I}(|E|) \}.$$

Also: Figalli, Maggi, Pratelli (2010) using ideas from OT.

Sharp Quantitative isoperimetric inequality in $\mathbb{R}^m$

Theorem (Fusco, Maggi, Pratelli 2008)

There exist constants $C, \delta > 0$ *such that for any* $|E| = |B^*|$ *and* $|E \ominus B^*| < \delta$ *one has*

$$\mathrm{Per}(E) - \mathrm{Per}(B^*) \geq C\alpha(E)^2.$$

Here, α *is the* **Fraenkel asymmetry***:*

$$\alpha(E) = \inf_{B^*}\{|E \ominus B^*| : |B^*| = |E|, \mathrm{Per}(B^*) = \mathbb{I}(|E|)\}.$$

Their proof doe not exhibit a “right notion of convexity", and instead relies on symmetrization arguments. However, the proof of **Cicalese and Leonardi 12'** does, and allows for extension to manifolds...

Quantitative isoperimetric inequality in $\mathcal{M}$

Theorem (Chodosh, Engelstein, Spolaor 19')

Suppose that a mass-constrained perimeter minimizer $E^* \subset \mathcal{M}$ *has* ***smooth*** *boundary* ∂E^**. Furthermore, suppose that either* E^* *is a* ***strict*** *perimeter minimizer, or that it is* ***integrable****. Then there exists constants* $C, \delta > 0$ *such that for any* $|E| = |E^*|$ *and* $|E \ominus E^*| < \delta$ *one has*

$$\mathrm{Per}(E) - \mathrm{Per}(E^*) \geq C\alpha(E)^2. \tag{1}$$

Main results

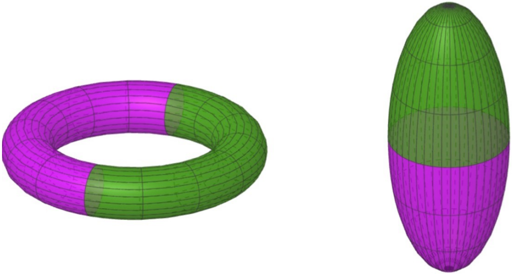
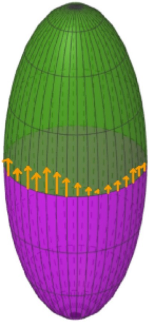


Figure: **Integrable** and **strict** minimizers

The “right" notion of convexity:



$\vec{V} = f\vec{\nu}$

$$D^2\,\mathrm{Per}(E^*)[f,f] \geq C\|f\|^2_{W^{1,2}(\partial E^*)}$$

Convergence of Cheeger cuts

Theorem:(NGT, Murray, Thorpe 20') Assume:

- The function $v \mapsto \frac{\mathbb{I}(v)}{\min\{v, |\mathcal{M}| - v\}}$ is locally strongly convex at its minimizers (denote this set by $\mathcal{V}^*$).
- Minimizers of isoperimetric problem for $v \in \mathcal{V}^*$ are smooth and either **integrable** or **strict**.

Then, as long as $\left(\frac{log(n)}{n}\right)^{1/m} \ll \varepsilon \ll 1$, w.v.h.p., for every minimizer A_n^* of graph Cheeger cut problem there exists A^* solution to the continuum Cheeger cut problem satisfying:

$$\|\mathbb{1}_{A_n^*} - \mathbb{1}_{A^*}\|_{L^1} \leq C\left(\frac{\log(n)^{1/m}}{n^{1/m}\varepsilon} + \sqrt[3]{\varepsilon}\right)^{1/8}.$$

Convergence of Cheeger cuts

Theorem:(NGT, Murray, Thorpe 20') Assume:

- The function $v \mapsto \frac{\mathbb{I}(v)}{\min\{v, |\mathcal{M}| - v\}}$ is locally strongly convex at its minimizers (denote this set by $\mathcal{V}^*$).
- Minimizers of isoperimetric problem for $v \in \mathcal{V}^*$ are smooth and either **integrable** or **strict**.

Then, as long as $\left(\frac{log(n)}{n}\right)^{1/m} \ll \varepsilon \ll 1$, w.v.h.p., for every minimizer A_n^* of graph Cheeger cut problem there exists A^* solution to the continuum Cheeger cut problem satisfying:

$$\|\mathbb{1}_{A_n^*} - \mathbb{1}_{A^*}\|_{L^1} \leq C \left(\frac{\log(n)^{1/m}}{n^{1/m}\varepsilon} + \sqrt[3]{\varepsilon} \right)^{1/8}.$$

Convergence of Cheeger cuts

Theorem:(NGT, Murray, Thorpe 20') Assume:

- The function $v \mapsto \frac{\mathbb{I}(v)}{\min\{v, |\mathcal{M}| - v\}}$ is locally strongly convex at its minimizers (denote this set by $\mathcal{V}^*$).
- Minimizers of isoperimetric problem for $v \in \mathcal{V}^*$ are smooth and either **integrable** or **strict**.

Then, as long as $\left(\frac{\log(n)}{n}\right)^{1/m} \ll \varepsilon \ll 1$, w.v.h.p., for every minimizer A_n^* of graph Cheeger cut problem there exists A^* solution to the continuum Cheeger cut problem satisfying:

$$\|\mathbb{1}_{A_n^*} - \mathbb{1}_{A^*}\|_{L^1} \leq C \left(\frac{\log(n)^{1/m}}{n^{1/m}\varepsilon} + \sqrt[3]{\varepsilon} \right)^{\cancel{1/8} \nearrow 1/2}.$$

What this talk was really about:

1. The use of geometry for learning, and the learning of geometry from data.
2. Passing the probabilistic estimates to where they don't harm.
3. Find the right notion of convexity.

What this talk was really about:

1. The use of geometry for learning, and the learning of geometry from data.
2. Passing the probabilistic estimates to where they don't harm.
3. Find the right notion of convexity.

From graph cuts to isoperimetric inequalities: Convergence rates of Cheeger cuts on data clouds (2020)

https://arxiv.org/abs/2004.09304

Thank you for your attention!