Regularity theory and uniform convergence in the large data limit of graph Laplacian eigenvectors on random data clouds

Nicolás García Trillos
University of Wisconsin-Madison
Joint work with Jeff Calder (UMinnesota) and Marta Lewicka (UPitt)

MSRI Workshop: Optimal transport and Applications to Machine Learning and Statistics
May 2020

Outline

- ACT 0: Introduction and pointwise convergence rates of graph Laplacians.
- ACT 1: Asymptotic spectral consistency.
- ACT 2: L^2-convergence rates.
- ACT 3: Regularity and Almost $\mathcal{C}^{0,1}$-convergence.

ACT 0: Introduction and pointwise convergence rates of graph Laplacians.

Goal: Given a data set $\mathcal{M}_n = \{x_1, \ldots, x_n\}$:

- Unsupervised learning: Find coarse structure of $\mathcal{M}_n$ (find meaningful clusters).
- Supervised learning: If in addition we have labels $y_1, \ldots, y_p$ associated to $x_1, \ldots, x_n$, find regression function $u : x \mapsto y$.

Goal: Given a data set $\mathcal{M}_n = \{x_1, \ldots, x_n\}$:

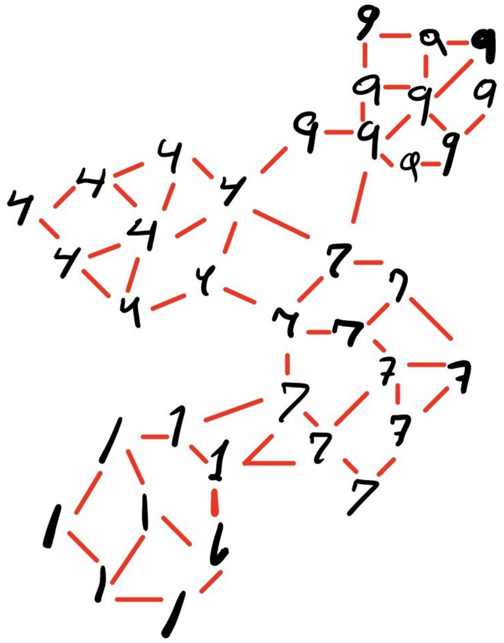
- Unsupervised learning: Find coarse structure of $\mathcal{M}_n$ (find meaningful clusters).
- Supervised learning: If in addition we have labels $y_1, \ldots, y_p$ associated to $x_1, \ldots, x_n$, find regression function $u : x \mapsto y$.

Graph based learning

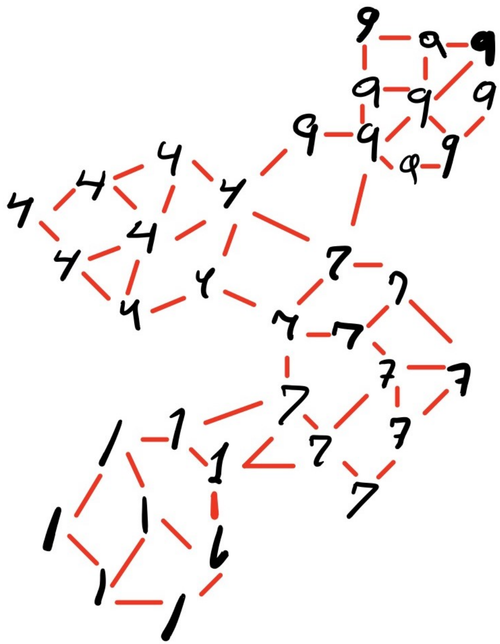
Goal: Given a data set $\mathcal{M}_n = \{x_1, \dots, x_n\}$ and similarity matrix $\{\omega_{ij}\}_{ij}$ do unsupervised/supervised learning.

Graph based learning

Given $G = (\mathcal{M}_n, \omega)$:



Given $G = (\mathcal{M}_n, \omega)$:



Graph Laplacian Δ_G:

$$\Delta_G u(x_i) := \sum_{ij} \omega_{ij}(u(x_i) - u(x_j)), \quad x_i \in \mathcal{M}_n.$$

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

$$x_i \in X_n \mapsto \begin{pmatrix} u_1(x_i) \\ \vdots \\ u_N(x_i) \end{pmatrix} \in \mathbb{R}^N$$

where $u_1, \dots, u_N$ first N eigenvectors of Δ_G.

Supervised Learning: Zhu et al (2003)

$$\text{argmin}_{u:X\to\mathbb{R}} \langle \Delta_G^\alpha u, u\rangle + L(y;u), \quad \text{e.g. } L(y;u) = \frac{1}{2\sigma^2}\sum_{i=1}^p |u(x_i) - y_i|^2$$

or Bayesian setting as in Zhu et al (2003), Kirichenko and van Zanten (2017), Bertozzi et al (2018):

$$y|u \sim \exp(-L(y;u)), \quad u \sim \pi = N(0, \Delta_\Gamma^{-\alpha})$$

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

$$x_i \in X_n \mapsto \begin{pmatrix} u_1(x_i) \\ \vdots \\ u_N(x_i) \end{pmatrix} \in \mathbb{R}^N$$

where $u_1, \dots, u_N$ first N eigenvectors of Δ_G.
Supervised Learning: Zhu et al (2003)

$$\mathrm{argmin}_{u:X\to\mathbb{R}} \langle \Delta_G^\alpha u, u\rangle + L(y;u), \qquad \text{e.g. } L(y;u) = \frac{1}{2\sigma^2}\sum_{i=1}^{p} |u(x_i) - y_i|^2$$

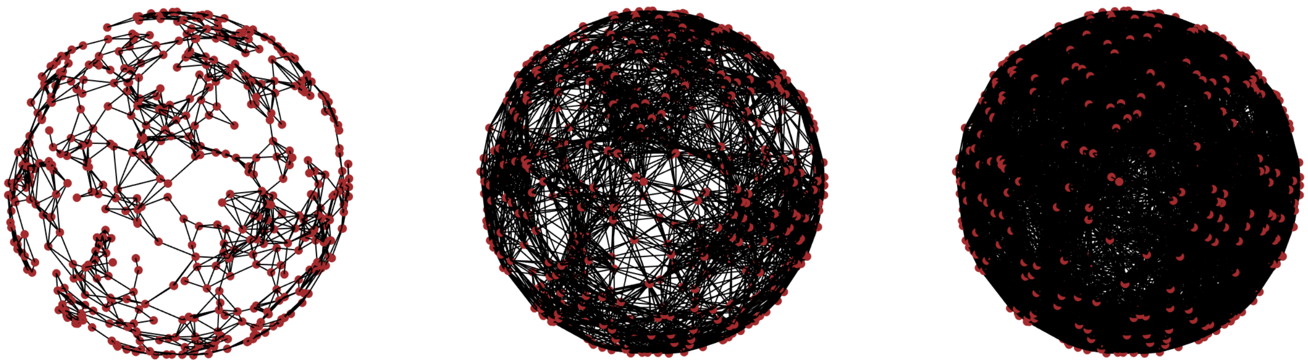
or Bayesian setting as in Zhu et al (2003), Kirichenko and van Zanten (2017), Bertozzi et al (2018):

$$y|u \sim \exp(-L(y;u)), \quad u \sim \pi = N(0, \Delta_\Gamma^{-\alpha})$$

Statistics of graph based methodologies under some modeling assumption?

Proximity graphs

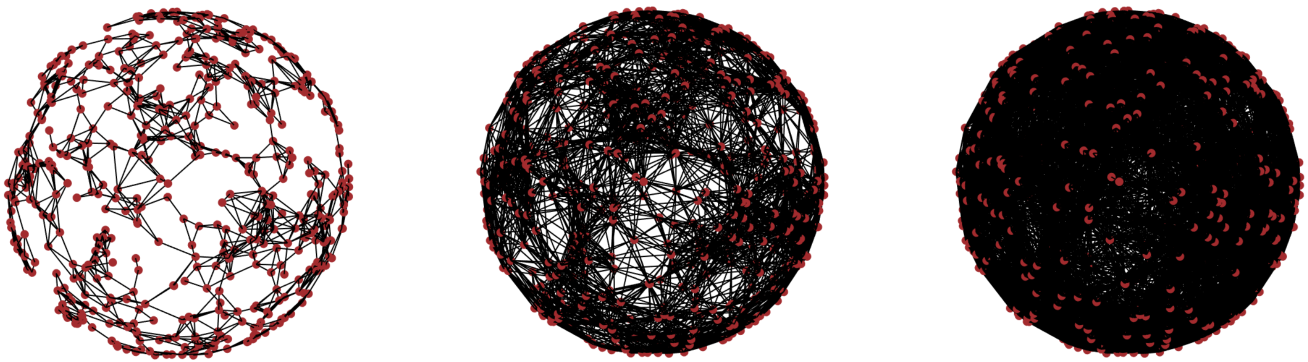
- $\mathcal{M}_n := \{x_1, \ldots, x_n\} \subseteq \mathcal{M} \subseteq \mathbb{R}^d$ with $m \ll d$ (**the manifold assumption**).



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

Proximity graphs

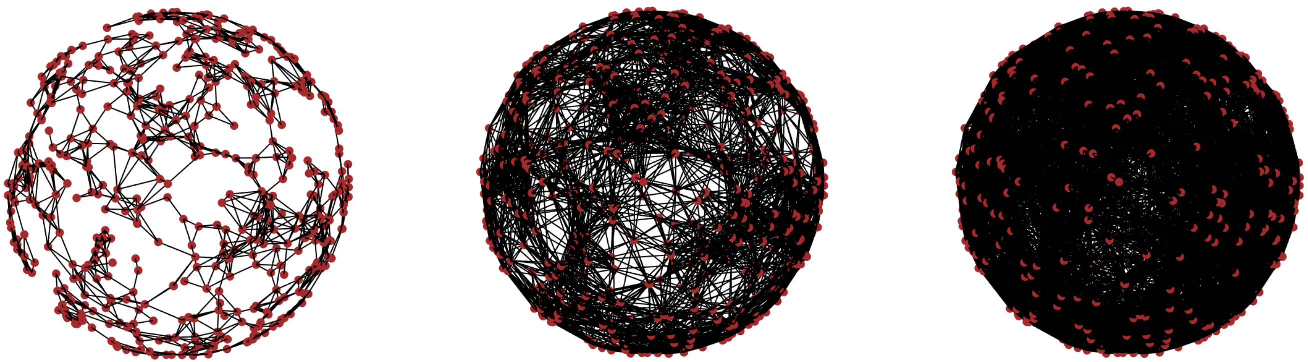
- $\mathcal{M}_n := \{x_1, \dots, x_n\} \subseteq \mathcal{M} \subseteq \mathbb{R}^d$ with $m \ll d$ (**the manifold assumption**).



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

Proximity graphs

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



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

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

Manifold Learning for **graph Laplacian**:

- Towards a theoretical foundation for Laplacian based methods. Belkin and Niyogi (2005).
- From graphs to manifolds: weak and strong poitwise consistency of graph Laplacians. Hein et al (2005).
- Diffusion maps. Coifman and Lafon (2005).
- Graph Laplacinans and their convergence on neihborhood graphs. Hein et al (2005).
- From graph to manifold Laplacian: the convergence rate. Singer (2006).
- ⋮

Pointwise Consistency

Theorem: [Hein et al, and others] $\mathcal{M}_n = \{x_1, \dots, x_n\}$ i.i.d. samples from distribution $d\mu(x) = \rho(x) d\,\mathrm{Vol}_{\mathcal{M}}(x)$.
Let $f \in C^3(\mathcal{M})$. Then, for $\varepsilon \leq \delta \leq \varepsilon^{-1}$:

$$\mathbb{P}\left[\max_{1 \leq i \leq n} |\Delta_\varepsilon f(x_i) - \Delta f(x_i)| \geq C\delta\right] \leq 2n \exp\left(-c\delta^2 n \varepsilon^{m+2}\right),$$

where C depends on $\|f\|_{C^3(\mathcal{M})}$.

Theorem: [Hein et al, and others] $\mathcal{M}_n = \{x_1, \ldots, x_n\}$ i.i.d. samples from distribution $d\mu(x) = \rho(x) d\,\mathrm{Vol}_{\mathcal{M}}(x)$.
Let $f \in C^3(\mathcal{M})$. Then, for $\varepsilon \leq \delta \leq \varepsilon^{-1}$:

$$\mathbb{P}\left[\max_{1\leq i\leq n} |\Delta_\varepsilon f(x_i) - \Delta f(x_i)| \geq C\delta\right] \leq 2n \exp\left(-c\delta^2 n\varepsilon^{m+2}\right),$$

where C depends on $\|f\|_{C^3(\mathcal{M})}$.
Here:

$$\Delta_{n,\varepsilon} f(x_i) := \frac{1}{n\varepsilon^2}\sum_{j=1}^{n} \eta_\varepsilon(|x_i - x_j|)(f(x_i) - f(x_j)),$$

$$\Delta f := -\frac{\sigma_\eta}{\rho}\mathrm{div}(\rho^2 \nabla f).$$

Pointwise Consistency

Theorem: [Hein et al, and others] $\mathcal{M}_n = \{x_1, \dots, x_n\}$ i.i.d. samples from distribution $d\mu(x) = \rho(x) d\,\mathrm{Vol}_{\mathcal{M}}(x)$.
Let $f \in C^3(\mathcal{M})$. Then, for $\varepsilon \leq \delta \leq \varepsilon^{-1}$:

$$\mathbb{P}\left[\max_{1 \leq i \leq n} |\Delta_\varepsilon f(x_i) - \Delta f(x_i)| \geq C\delta\right] \leq 2n \exp\left(-c\delta^2 n \varepsilon^{m+2}\right),$$

where C depends on $\|f\|_{C^3(\mathcal{M})}$.
Here:

$$\eta_\varepsilon(t) := \frac{1}{\varepsilon^m}\eta\left(\frac{t}{\varepsilon}\right)$$

$$\Delta_{n,\varepsilon} f(x_i) := \frac{1}{n\varepsilon^2}\sum_{j=1}^n \eta_\varepsilon(|x_i - x_j|)(f(x_i) - f(x_j)),$$

$$\Delta f := -\frac{\sigma_\eta}{\rho}\mathrm{div}(\rho^2 \nabla f).$$

ACT 1: Asymptotic Spectral Consistency

Consistency question

What happens as $n \to \infty$ with eigenvalues/eigenvectors of $\Delta_{n,\varepsilon}$?

- Consistency of Spectral clustering. von Luxburg, Belkin, and Bousquet (2007).
- A variational approach to the consistency of spectral clustering. NGT and Slepčev (2015).

Asymptotic spectral convergence: from $\Delta_{n,\varepsilon}$ to Δ

Theorem: [NGT, Slepčev (2015)] 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, for every $k \in \mathbb{N}$

$$\lim_{n \to \infty} \lambda^k_{n,\varepsilon} = \lambda^k, \quad \text{and } u_k \to_{TL^2} v_k.$$

Recall:

$$\Delta v(x) = -\frac{\sigma_\eta}{\rho} \mathrm{div}(\rho^2 \nabla v)$$

Due to Courant-Fisher min-max principle, we may study **minima/minimizers** of weighted **Dirichlet forms**:

Graph:

- Dirichlet energy:

$$D_{n,\varepsilon}(u) = \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\eta_\varepsilon(|x_i - x_j|)|u(x_i) - u(x_j)|^2$$

- Laplacian:

$$\Delta_{n,\varepsilon}u(x_i) = \frac{1}{n\varepsilon^2}\sum_{j=1}^{n}\eta_\varepsilon(|x_i - x_j|)(u(x_i) - u(x_j))$$

Continuum Local:

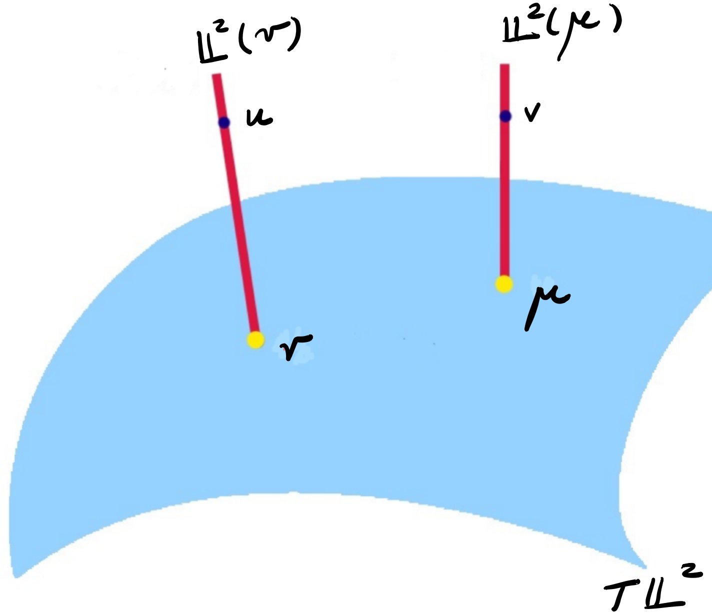
- Dirichlet energy:

$$D(v) = \sigma_\eta \int_{\mathcal{M}} |\nabla v|^2 \rho^2(x) d\,\mathrm{Vol}_{\mathcal{M}}(x)$$

- Laplacian:

$$\Delta v(x) = -\frac{1}{\rho}\mathrm{div}(\rho^2 \nabla v)$$

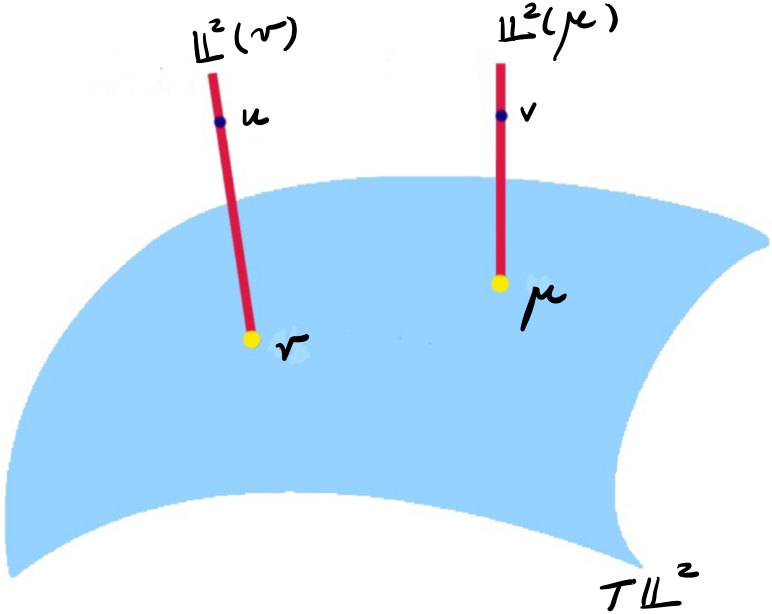
The TL^2 space:



$$TL^2 := \{(\nu, u) \ : \ \nu \in \mathcal{P}_2(\mathcal{M}), \quad u \in L^2(\nu)\}$$

$$\inf_{\pi \in \Gamma(\nu,\mu)} \int_{\mathcal{M}\times\mathcal{M}} \left(d_{\mathcal{M}}(x,y)^2 + |u(x) - v(y)|^2 \right) d\pi(x,y)$$

The TL^2 space:

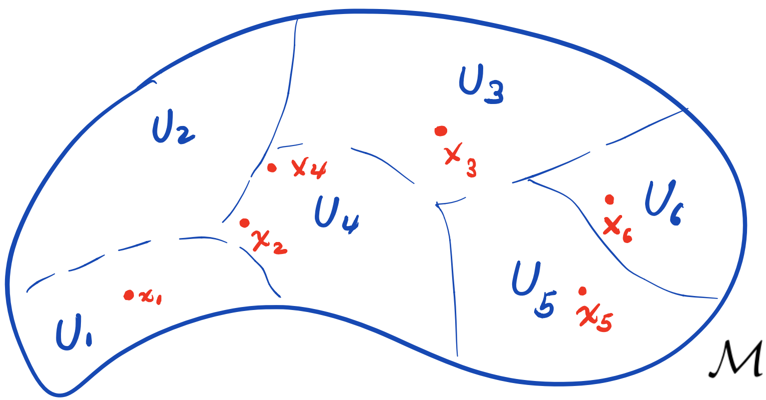


- Detailed discussion in: A Transportation L^p distance for signal analysis. Thorpe et al (2018).

Implicit interpolation

Let μ_n empirical measure of samples from μ. Let $u : \mathcal{M}_n \to \mathbb{R}$. $T_n : \mathcal{M} \to \mathcal{M}_n$ satisfying $T_\sharp \mu = \mu_n$ induces:

$$u \circ T_n : \mathcal{M} \to \mathbb{R}$$



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

Discrete to non-local control (use of interpolation map) + Analysis of Non-local Dirichlet energy.

Discrete to non-local control (use of interpolation map) + Analysis of Non-local Dirichlet energy.

Continuum Non-local:

- Dirichlet energy:

$$D_\varepsilon(v) = \frac{1}{\varepsilon^2}\int_{\mathcal{M}}\int_{\mathcal{M}} \eta_\varepsilon(|x-y|)|v(x)-v(y)|^2 d\mu(x)d\mu(y)$$

- Laplacian:

$$\Delta_\varepsilon v(x) = \frac{1}{\varepsilon^2}\int_{\mathcal{M}} \eta_\varepsilon(|x-y|)(v(x)-v(y))d\mu(y)$$

Discrete to Non-local (Interpolation Version 1)

$$\mathcal{I}_n^1 : L^2(\mu_n) \to L^2(\mu)$$

$u : X_n \to \mathbb{R}$ is mapped to

$$\mathcal{I}_n^1 u(x) = u \circ T_n.$$

This map satisfies:

1. $\|u\|_{L^2(\mu_n)} = \|u\|_{L^2(\mu_n)}$
2. $\left(1 - c\frac{\|Id - T_n\|_\infty}{\varepsilon}\right) D_{\tilde{\varepsilon}}(\mathcal{I}_n u) \leq D_{n,\varepsilon}(u)$, where $\tilde{\varepsilon} := \varepsilon - 2\|Id - T_n\|_\infty$.

Indeed, for a given **transport map** $T_{n\sharp}\mu = \mu_n$:

$$
\begin{aligned}
D_{n,\varepsilon}(u) &= \frac{1}{\varepsilon^{m+2}} \int\int \eta\left(\frac{|x-y|}{\varepsilon}\right)(u(x)-u(y))^2 d\mu_n(x) d\mu_n(y) \\
&= \frac{1}{\varepsilon^{m+2}} \int\int \eta\left(\frac{|T_n(x)-T_n(y)|}{\varepsilon}\right)(u(T_n(x))-u(T_n(y)))^2 d\mu(x) d\mu(y) \\
&\geq \frac{1}{\varepsilon^{m+2}} \int\int \eta\left(\frac{|x-y|}{\tilde{\varepsilon}}\right)(u(T_n(x))-u(T_n(y)))^2 d\mu(x) d\mu(y) \\
&\geq \left(1 - c\frac{\|Id - T_n\|_\infty}{\varepsilon}\right) D_{\tilde{\varepsilon}}(\mathcal{I}_n^1 u)
\end{aligned}
$$

$$
\tilde{\varepsilon} := \varepsilon - 2\|Id - T_n\|_\infty.
$$

The sequence $\{D_\varepsilon\}_\varepsilon$ Γ-converges in the $L^2(\mu)$ sense towards D as $\varepsilon \to 0$

Recall:

$$D_\varepsilon(v) = \frac{1}{\varepsilon^2} \int_{\mathcal{M}} \int_{\mathcal{M}} \eta_\varepsilon(|x-y|)|v(x) - v(y)|^2 d\mu(x) d\mu(y)$$

$$D(v) = \sigma_\eta \int_{\mathcal{M}} |\nabla v|^2 \rho^2(x) d\,\mathrm{Vol}_{\mathcal{M}}(x)$$

De Giorgi Γ-convergence: Sufficient (and for all purposes necessary) conditions for convergence of minimizers.

Non-local analysis

The sequence $\{D_\varepsilon\}_\varepsilon$ Γ-converges in the $L^2(\mu)$ sense towards D as $\varepsilon \to 0$

Recall:

$$D_\varepsilon(v) = \frac{1}{\varepsilon^2} \int_{\mathcal{M}} \int_{\mathcal{M}} \eta_\varepsilon(|x-y|)|v(x) - v(y)|^2 d\mu(x) d\mu(y)$$

$$D(v) = \sigma_\eta \int_{\mathcal{M}} |\nabla v|^2 \rho^2(x) d\,\mathrm{Vol}_{\mathcal{M}}(x)$$

- $\forall v_\varepsilon \to_{L^2(\mu)} v$: $D(v) \leq \liminf_{\varepsilon \to 0} D_\varepsilon(v_\varepsilon)$.
- $\forall v \; \exists v_\varepsilon \to_{L^2(\mu)} v$: $D(v) \geq \limsup_{\varepsilon \to 0} D_\varepsilon(v_\varepsilon)$.
- Every $\{v_\varepsilon\}_{\varepsilon > 0}$ with $\sup_{\varepsilon > 0} D_\varepsilon(v_\varepsilon) < \infty$, is precompact.

Combining: Discrete to Non-local + Analysis Non-local

It follows: $D_{n,\varepsilon}$ Γ-converges in TL^2 towards D, as $n \to \infty$, provided:

$$\inf_{T_\sharp \mu = \mu_n} \| T_n - Id \|_\infty \ll \varepsilon \ll 1$$

Combining: Discrete to Non-local + Analysis Non-local

It follows $D_{n,\varepsilon}$ Γ-converges in TL^2 towards D , as $n \to \infty$, provided:

$$d_\infty(\mu_n, \mu) = \inf_{T_{n\sharp}\mu = \mu_n} \| T_n - Id \|_\infty \ll \varepsilon \ll 1$$

If $\mathcal{M}_n = \{x_1, \dots, x_n\}$ are i.i.d. samples from μ

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

(Ajtai, Komlós, Tusnády (1984), Shor and Yukich (1989), Leighton and Shor (1993), Talagrand, NGT and Slepčev 2014).

ACT2: L^2 convergence rates

(Discrete to Non-local control + Analysis Non-local Dirichlet energy)

- A graph-discretization of the Laplace-Beltrami operator. Burago et al (2014).
- Error estimates for spectral convergence of the graph Laplacian on random geometric graphs towards the Laplace-Beltrami operator. NGT, Gerlach, Hein, Slepčev (2018).
- Improved spectral convergence rates for graph Laplacians on epsilon and k-NN graphs. Calder, NGT (2019).

Discrete to Non-local control (same as in the previous Act) +
Analysis Non-local Dirichlet energy (but more carefully) .

Analysis Non-local Dirichlet energy

Using ideas from Burago et al (2014):

$$D(\Lambda_\varepsilon v) \leq (1 + C\varepsilon) D_\varepsilon(v), \quad \forall v \in L^2(\mu)$$

where

$$\Lambda_r v(x) \propto \int_{\mathcal{M}} \psi_r(d_{\mathcal{M}}(x, y)) v(y) d\mu(y), \quad \psi(t) := \frac{1}{\sigma_\eta} \int_t^\infty \eta(s) s ds$$

L^2 convergence rates: version 1

Theorem:[NGT, Gerlach, Hein, Slepčev, 2019] As long as $\left(\frac{log(n)}{n}\right)^{p_m} \ll \varepsilon \ll 1$, w.v.h.p.,

$$|\lambda^k_{n,\varepsilon} - \lambda_k| \leq C\left(\frac{d_\infty(\mu_n, \mu)}{\varepsilon} + (\sqrt{\lambda^k} + 1)\varepsilon\right)\lambda^k$$

$$\|\mathcal{I}^1_n u^k - v_k\|^2_{L^2(\mu)} \leq C_k\left(\frac{d_\infty(\mu_n, \mu)}{\varepsilon} + (\sqrt{\lambda^k} + 1)\varepsilon\right)\lambda^k$$

Recall: ∞-OT distance between μ and μ_n,

$$d_\infty(\mu, \mu_n) \sim \frac{(\log(n))^{p_m}}{n^{1/m}}.$$

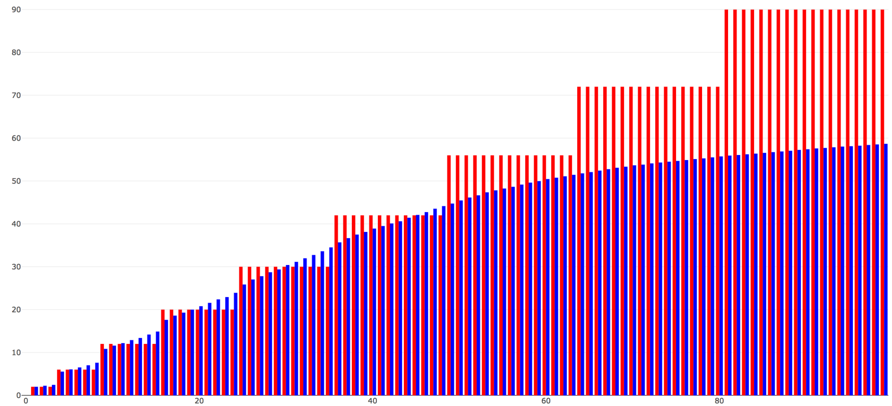


Figure: Eigenvalues Graph Laplacian Vs Eigenvalues Laplacian on Sphere.

$$|\lambda_{n,\varepsilon}^k - \lambda_k| \leq C \left(\frac{d_\infty(\mu_n, \mu)}{\varepsilon} + (\sqrt{\lambda^k} + 1)\varepsilon \right) \lambda^k$$

Recall: ∞-transportation distance between μ and μ_n,

$$d_\infty(\mu, \mu_n) \sim \frac{(\log(n))^{p_m}}{n^{1/m}}.$$

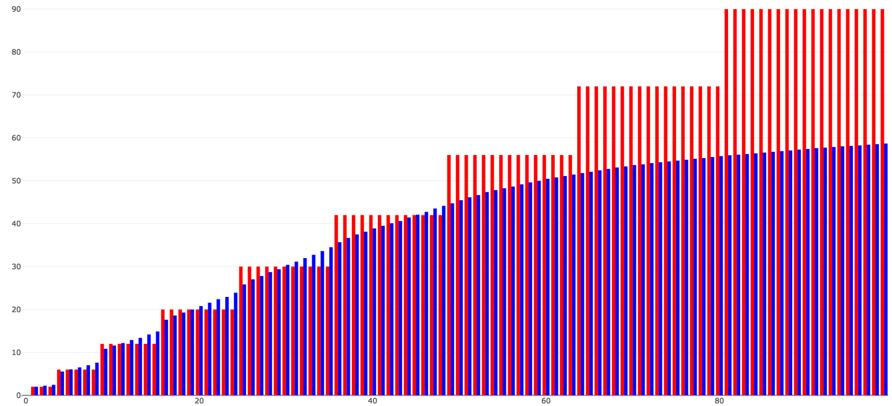


Figure: Eigenvalues Graph Laplacian Vs Eigenvalues Laplacian on Sphere.

$$|\lambda_{n,\varepsilon}^k - \lambda_k| \leq C \left(\frac{d_\infty(\mu_n, \mu)}{\varepsilon} + (\sqrt{\lambda^k} + 1)\varepsilon \right) \lambda^k$$

Recall: ∞-transportation distance between μ and μ_n,

$$d_\infty(\mu, \mu_n) \sim \frac{(\log(n))^{p_m}}{n^{1/m}}.$$

Side Remark: From graph cuts to isoperimetric inequalities: Convergence rates of Cheeger cuts on data clouds (NGT, Murray, Thorpe 2020)


INFINITYTRANSPORTATION
312.541.2222

Improving the convergence rates

Suppose $\Delta_{n,\varepsilon} u = \lambda_{n,\varepsilon} u$ and $\Delta v = \lambda v$ both normalized.

Improving the convergence rates

Suppose $\Delta_{n,\varepsilon} u = \lambda_{n,\varepsilon} u$ and $\Delta v = \lambda v$ both normalized.

$$\begin{aligned}
\lambda_{n,\varepsilon} \langle u, v \rangle_{L^2(\mu_n)} &= \langle \Delta_{n,\varepsilon} u, v \rangle_{L^2(\mu_n)} \\
&= \langle u, \Delta_{n,\varepsilon} v \rangle_{L^2(\mu_n)} \\
&= \langle u, \Delta v \rangle_{L^2(\mu_n)} + \langle u, \Delta_{n,\varepsilon} v - \Delta v \rangle_{L^2(\mu_n)} \\
&= \lambda \langle u, v \rangle_{L^2(\mu_n)} + \langle u, \Delta_{n,\varepsilon} v - \Delta v \rangle_{L^2(\mu_n)}
\end{aligned}$$

The bottom line is:

$$|\lambda - \lambda_{n,\varepsilon}| \leq \frac{\max_{i=1,\dots,n} |\Delta v(x_i) - \Delta_{n,\varepsilon} v(x_i)|}{|\langle u, v \rangle_{L^2(\mu_n)}|}$$

Improving the convergence rates

Suppose $\Delta_{n,\varepsilon} u = \lambda_{n,\varepsilon} u$ and $\Delta v = \lambda v$ both normalized.

$$
\begin{aligned}
\lambda_{n,\varepsilon} \langle u, v \rangle_{L^2(\mu_n)} &= \langle \Delta_{n,\varepsilon} u, v \rangle_{L^2(\mu_n)} \\
&= \langle u, \Delta_{n,\varepsilon} v \rangle_{L^2(\mu_n)} \\
&= \langle u, \Delta v \rangle_{L^2(\mu_n)} + \langle u, \Delta_{n,\varepsilon} v - \Delta v \rangle_{L^2(\mu_n)} \\
&= \lambda \langle u, v \rangle_{L^2(\mu_n)} + \langle u, \Delta_{n,\varepsilon} v - \Delta v \rangle_{L^2(\mu_n)}
\end{aligned}
$$

The bottom line is:

$$
|\lambda - \lambda_{n,\varepsilon}| \leq \frac{\max_{i=1,\dots,n} |\Delta v(x_i) - \Delta_{n,\varepsilon} v(x_i)|}{|\langle u, v \rangle_{L^2(\mu_n)}|}
$$

In other words: with a priori estimates we can upgrade to the pointwise convergence rates .

Improved L^2-convergence rates

Theorem:[Calder, NGT, 2019] Let $k \in \mathbb{N}$. As long as

$$\left(\frac{\log(n)}{n}\right)^{1/(m+4)} \ll \varepsilon \ll 1,$$

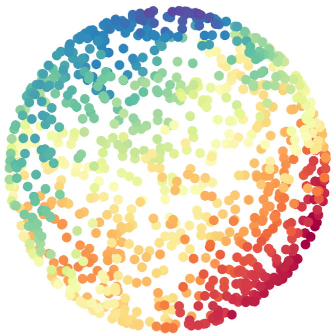
then w.v.h.p:

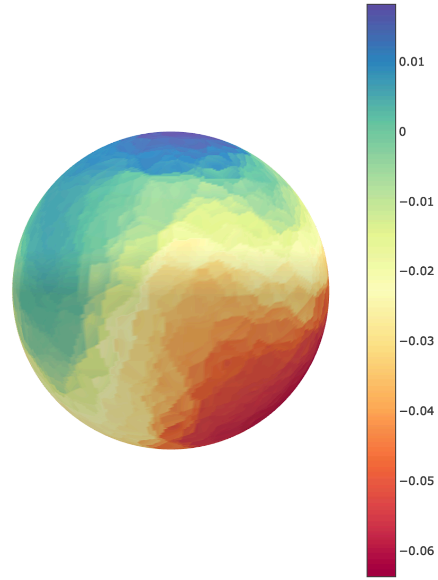
$$|\lambda_{n,\varepsilon}^k - \lambda^k| \leq C_k \varepsilon$$

and

$$\|\mathcal{I}_n^1 u_k - v_k\|_{L^2(\mu)} \leq C_k \varepsilon, \quad \|u_k - v_k\|_{L^2(\mu_n)} \leq C_k \varepsilon.$$

ACT3: Regularity and uniform convergence results.




0.01
0
−0.01
−0.02
−0.03
−0.04
−0.05
−0.06

Global Lipschitz estimates

Theorem: [Calder, NGT, Lewicka, 2020] Let ε be small enough. As long as

$$\left(\frac{\log(n)}{n}\right)^{1/(m+4)} \ll \varepsilon \ll 1,$$

then w.v.h.p:

$$|\tilde{u}(x_i) - \tilde{u}(x_j)| \leq C\big(\|\tilde{u}\|_{L^\infty(\mathcal{M}_n)} + \|\Delta_{n,\varepsilon}\tilde{u}\|_{L^\infty(\mathcal{M}_n)}\big) \cdot (d_{\mathcal{M}}(x_i, x_j) + \varepsilon)$$

for all $\tilde{u} : \mathcal{M}_n \to \mathbb{R}$ and all $x_i, x_j \in \mathcal{M}_n$.

Corollary 1: Regularity of eigenvectors

The general inequality is:

$$|\tilde{u}(x_i) - \tilde{u}(x_j)| \leq C\left(\|\tilde{u}\|_{L^\infty(\mathcal{M}_n)} + \|\Delta_{n,\varepsilon}\tilde{u}\|_{L^\infty(\mathcal{M}_n)}\right) \cdot \left(d_{\mathcal{M}}(x_i, x_j) + \varepsilon\right).$$

Corollary 1: Regularity of eigenvectors

The general inequality is:

$$|\tilde{u}(x_i) - \tilde{u}(x_j)| \leq C\left(\|\tilde{u}\|_{L^\infty(\mathcal{M}_n)} + \|\Delta_{n,\varepsilon}\tilde{u}\|_{L^\infty(\mathcal{M}_n)}\right) \cdot (d_{\mathcal{M}}(x_i, x_j) + \varepsilon).$$

Take $\tilde{u} = u$ where: $\Delta_{n,\varepsilon} u = \lambda_{n,\varepsilon} u$ and $\|u\|_{L^2(\mu_n)} = 1$. Then,

$$|u(x_i) - u(x_j)| \leq C\left((\lambda_{n,\varepsilon} + 1)\|u\|_{L^\infty(\mathcal{M}_n)}\right) \cdot (d_{\mathcal{M}}(x_i, x_j) + \varepsilon)$$

In fact, from the above it follows:

$$\|u\|_{L^\infty(X_n)} \leq C(\lambda_{n,\varepsilon} + 1)^m \|u\|_{L^1(X_n)},$$

provided $\varepsilon \leq \frac{c}{\lambda_{n,\varepsilon}+1}$. Hence...

Corollary 1: Regularity of eigenvectors

Bottom line: $\Delta_{n,\varepsilon} u = \lambda_{n,\varepsilon} u$ and $\|u\|_{L^2(\mu_n)} = 1$. Then,

$$|u(x_i) - u(x_j)| \leq C((\lambda_{n,\varepsilon} + 1)^{m+1}) \cdot (d_{\mathcal{M}}(x_i, x_j) + \varepsilon)$$

or

$$[u]_{1,\varepsilon} := \max_{x_i \neq x_j} \frac{|u(x_i) - u(x_j)|}{d_{\mathcal{M}}(x_i, x_j) + \varepsilon} \leq C(\lambda_{n,\varepsilon} + 1)^{m+1}.$$

i.e. u is "almost" Lipschitz.

Corollary 2: Almost $\mathcal{C}^{0,1}$-convergence

The general inequality is:

$$|\tilde{u}(x_i) - \tilde{u}(x_j)| \leq C\big(\|\tilde{u}\|_{L^\infty(\mathcal{M}_n)} + \|\Delta_{n,\varepsilon}\tilde{u}\|_{L^\infty(\mathcal{M}_n)}\big) \cdot (d_{\mathcal{M}}(x_i, x_j) + \varepsilon).$$

Corollary 2: Almost $\mathcal{C}^{0,1}$ convergence rates

The general inequality is:

$$|\tilde{u}(x_i) - \tilde{u}(x_j)| \leq C(\|\tilde{u}\|_{L^\infty(\mathcal{M}_n)} + \|\Delta_{n,\varepsilon} u\|_{L^\infty(\mathcal{M}_n)}) \cdot (d_{\mathcal{M}}(x_i, x_j) + \varepsilon)$$

Take $\tilde{u} = u - v$ where $\Delta_n u = \lambda_{n,\varepsilon} u$ and $\Delta v = \lambda v$ both normalized.

Note:

$$\Delta_{n,\varepsilon}(u - v) = \lambda_{n,\varepsilon}(u - v) + (\lambda - \lambda_{n,\varepsilon})v + (\Delta_{n,\varepsilon} v - \Delta v)$$

We can essentially obtain:

$$\|u - v\|_{L^\infty(\mathcal{M}_n)} \leq C(\lambda + 1)^m \|u - v\|_{L^2(\mathcal{M}_n)},$$

and then

$$[u - v]_{1,\varepsilon} \leq C(\lambda + 1)^{m+1} \|u - v\|_{L^2(\mathcal{M}_n)}.$$

How do we prove this?

Theorem[Calder, NGT, Lewicka, 2020]: Let ε be small enough. As long as

$$\left(\frac{\log(n)}{n}\right)^{1/(m+4)} \ll \varepsilon \ll 1,$$

then w.v.h.p:

$$|\tilde{u}(x_i) - \tilde{u}(x_j)| \leq C\left(\|\tilde{u}\|_{L^\infty(\mathcal{M}_n)} + \|\Delta_{n,\varepsilon}\tilde{u}\|_{L^\infty(\mathcal{M}_n)}\right) \cdot \left(d_{\mathcal{M}}(x_i, x_j) + \varepsilon\right)$$

for all $\tilde{u} \in L^2(\mathcal{M}_n)$ and all $x_i, x_j \in \mathcal{M}_n$.

Discrete to Non-local control (with different interpolator) + Analysis Non-local Laplacian.

Step 1: Discrete to non-local control

Interpolation map (Version 2): Given $u : \mathcal{M}_n \to \mathbb{R}$ define $\mathcal{I}^2_{n,\varepsilon} u : \mathcal{M} \to \mathbb{R}$

$$\mathcal{I}^2_{n,\varepsilon} u(x) := \frac{1}{n d_{\varepsilon,n}(x)} \sum_{i=1}^{n} \eta_\varepsilon(|x - x_i|) u(x_i).$$

Step 1: Discrete to non-local control

$$\mathcal{I}^2_{n,\varepsilon} u(x) := \frac{1}{n d_{\varepsilon,n}(x)} \sum_{i=1}^{n} \eta_\varepsilon(|x - x_i|) u(x_i).$$

Properties: Let ε be small enough. As long as

$$\left(\frac{\log(n)}{n}\right)^{1/(m+4)} \ll \varepsilon \ll 1,$$

then w.v.h.p:

1. $|u(x_i) - u(x_j)| \leq C\varepsilon^2 \|\Delta_{n,\varepsilon} u\|_{L^\infty(\mathcal{M}_n)} + |\mathcal{I}^2_{n,\varepsilon} u(x_i) - \mathcal{I}^2_{n,\varepsilon} u(x_j)|$
2. $\|\mathcal{I}^2_{n,\varepsilon} u\|_{L^\infty(\mathcal{M})} \leq C \|u\|_{L^\infty(\mathcal{M}_n)}$
3. $\|\Delta_\varepsilon(\mathcal{I}^2_{n,\varepsilon} u)\|_{L^\infty(\mathcal{M})} \leq C\left(\|\Delta_{n,\varepsilon} u\|_{L^\infty(\mathcal{M}_n)} + \|u\|_{L^\infty(\mathcal{M}_n)}\right).$

Step 1: Discrete to non-local control

$$\mathcal{I}^2_{n,\varepsilon} u(x) := \frac{1}{n d_{\varepsilon,n}(x)} \sum_{i=1}^{n} \eta_\varepsilon(|x - x_i|) u(x_i).$$

Properties: Let ε be small enough. As long as

$$\left(\frac{\log(n)}{n}\right)^{1/(m+4)} \ll \varepsilon \ll 1,$$

then w.v.h.p:

1. $|u(x_i) - u(x_j)| \leq C\varepsilon^2 \|\Delta_{n,\varepsilon} u\|_{L^\infty(\mathcal{M}_n)} + |\mathcal{I}^2_{n,\varepsilon} u(x_i) - \mathcal{I}^2_{n,\varepsilon} u(x_j)|$
2. $\|\mathcal{I}^2_{n,\varepsilon} u\|_{L^\infty(\mathcal{M})} \leq C\|u\|_{L^\infty(\mathcal{M}_n)}$
3. $\|\Delta_\varepsilon(\mathcal{I}^2_{n,\varepsilon} u)\|_{L^\infty(\mathcal{M})} \leq C\left(\|\Delta_{n,\varepsilon} u\|_{L^\infty(\mathcal{M}_n)} + \|u\|_{L^\infty(\mathcal{M}_n)}\right).$

Recall:

$$\Delta_\varepsilon v(x) = \frac{1}{\varepsilon^2} \int_{\mathcal{M}} \eta_\varepsilon(|x - y|)(v(x) - v(y)) d\mu(y)$$

Step 2: Analysis Non-local Laplacian

Theorem:(Global regularity Non-local Laplacian) Let ε be small enough. Then, for every $v : \mathcal{M} \to \mathbb{R}$

$$|v(x) - v(y)| \leq C\big(\|v\|_{L^\infty(\mathcal{M})} + \|\Delta_\varepsilon v\|_{L^\infty(\mathcal{M})}\big) \cdot \big(d_{\mathcal{M}}(x, y) + \varepsilon\big),$$

for all $x, y \in \mathcal{M}$.

Step 2: Analysis Non-local Laplacian

Theorem:(Global regularity Non-local Laplacian) Let ε be small enough. Then, for every $v : \mathcal{M} \to \mathbb{R}$

$$|v(x) - v(y)| \leq C\left(\|v\|_{L^\infty(\mathcal{M})} + \|\Delta_\varepsilon v\|_{L^\infty(\mathcal{M})}\right) \cdot (d_{\mathcal{M}}(x,y) + \varepsilon),$$

for all $x, y \in \mathcal{M}$.

Coupling of random walks methods to obtain regularity estimates:

- Cranston (1991): using coupling of Brownian particles proposed by Lindvall and Rogers (1986).
- F.-Y. Wang (1994, 2004).
- Priola and Wang (2006).
- Kusuoka (2017).
- Porretta and Priola (2013).

Future Work

- Convergence rates using a priori regularity and classical Statistical Learning tools.
- Analysis in semisupervised learning with fractional Laplacian regularization in Bayesian and optimization settings.
  - Continuum limits of posteriors in graph bayesian inverse problems. NGT, Sanz-Alonso (SIMA, 2018).
  - On the consistency of graph-based Bayesian learning and the scalability of sampling algorithms. NGT, Kaplan, Samakhoana, Sanz-Alonso (JMLR, 2020).
- Numerical analysis of PDEs on manifolds: Meshless methods with unstructured point clouds.

Thank you for your attention!

Special thanks to:
- NSF: NSF Grant DMS-1912818.
-All my collaborators.