Discrete regularity for graph Laplacians

Jeff Calder

School of Mathematics
University of Minnesota

Workshop on Stochastic Analysis Related to Hamilton-Jacobi PDEs
Institute for Pure and Applied Mathematics

May 19, 2020

Joint work with: Nicolas Garcia Trillos (Wisconsin) and Marta Lewicka (Pittsburgh)

Research supported by NSF-DMS grant 1713691

Outline

Outline

Outline

Graph-based learning

Let $(\mathcal{X}, \mathcal{W})$ be a graph.

- $\mathcal{X} \subset \mathbb{R}^d$ are the vertices.
- $\mathcal{W} = (w_{xy})_{x,y\in\mathcal{X}}$ are nonnegative edge weights.

Graph-based learning

Let $(\mathcal{X}, \mathcal{W})$ be a graph.

- $\mathcal{X} \subset \mathbb{R}^d$ are the vertices.
- $\mathcal{W} = (w_{xy})_{x,y \in \mathcal{X}}$ are nonnegative edge weights.

In data science/machine learning, data is often given a graph structure. In this case w_{xy} is large when x and y are similar, and small or $w_{xy} = 0$ otherwise.

Graph-based learning

Let $(\mathcal{X}, \mathcal{W})$ be a graph.

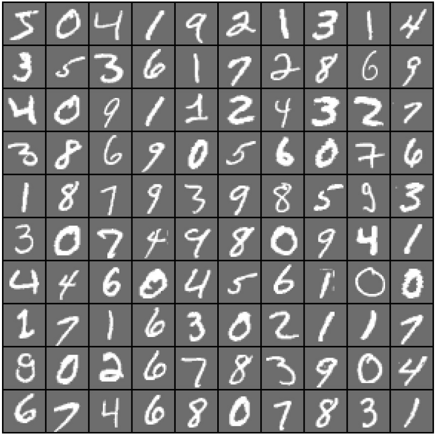
- $\mathcal{X} \subset \mathbb{R}^d$ are the vertices.
- $\mathcal{W} = (w_{xy})_{x,y\in\mathcal{X}}$ are nonnegative edge weights.

In data science/machine learning, data is often given a graph structure. In this case w_{xy} is large when x and y are similar, and small or $w_{xy} = 0$ otherwise.

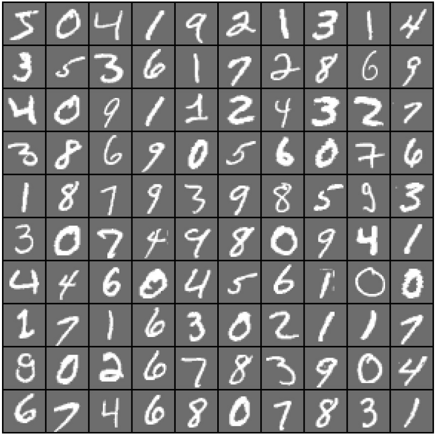
Common graph-based learning tasks

- Clustering
  - Grouping similar datapoints
- Semi-supervised learning.
  - Clustering with some label information.

MNIST (70,000 28×28 pixel images of digits 0-9)



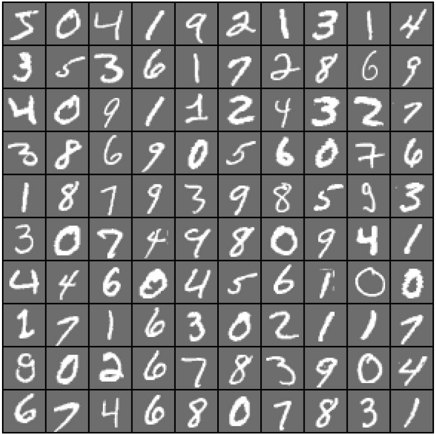
MNIST (70,000 28×28 pixel images of digits 0-9)



- Each image is a datapoint

$$x \in \mathbb{R}^{28\times 28} = \mathbb{R}^{784}.$$

MNIST (70,000 28×28 pixel images of digits 0-9)



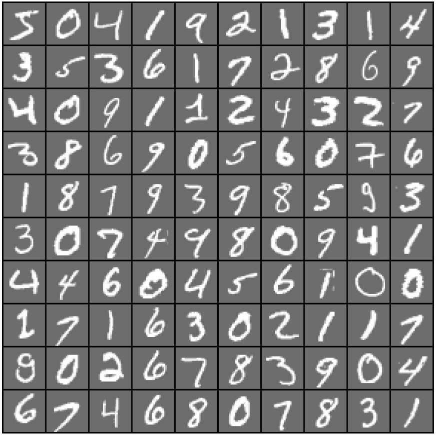
- Each image is a datapoint

$$x \in \mathbb{R}^{28\times 28} = \mathbb{R}^{784}.$$

- Geometric weights:

$$w_{xy} = \eta\left(\frac{|x-y|}{\varepsilon}\right)$$

MNIST (70,000 28×28 pixel images of digits 0-9)



- Each image is a datapoint

$$x \in \mathbb{R}^{28\times 28} = \mathbb{R}^{784}.$$

- Geometric weights:

$$w_{xy} = \eta\left(\frac{|x-y|}{\varepsilon}\right)$$

- k-nearest neighbor graph:

$$w_{xy} = \eta\left(\frac{|x-y|}{\varepsilon_k(x)}\right)$$

Clustering MNIST

https://divamgupta.com

Outline

Graph cuts

Question: How do we cluster graph data?

Graph cuts

Question: How do we cluster graph data?

Consider binary clustering (two classes). We can try to minimize a graph cut energy

$$\text{(Min-Cut)} \qquad \min_{A \subset \mathcal{X}} \mathrm{Cut}(A) := \sum_{\substack{x,y \in \mathcal{X} \\ x \in A, y \notin A}} w_{xy}.$$

Graph cuts

Question: How do we cluster graph data?

Consider binary clustering (two classes). We can try to minimize a graph cut energy

$$\text{(Min-Cut)} \qquad \min_{A \subset \mathcal{X}} \operatorname{Cut}(A) := \sum_{\substack{x,y \in \mathcal{X} \\ x \in A,\, y \notin A}} w_{xy}.$$

Tends to produce unbalanced classes (e.g., $A = \{x\}$).

Graph cuts

Question: How do we cluster graph data?

Consider binary clustering (two classes). We can try to minimize a graph cut energy

$$\text{(Balanced-Cut)} \qquad \min_{A \subset \mathcal{X}} \frac{\mathrm{Cut}(A)}{\mathrm{Vol}(A)\mathrm{Vol}(\mathcal{X} \setminus A)},$$

where

$$\mathrm{Vol}(A) = \sum_{x \in A} \sum_{y \in X} w_{xy}.$$

Graph cuts

Question: How do we cluster graph data?

Consider binary clustering (two classes). We can try to minimize a graph cut energy

$$\text{(Balanced-Cut)} \qquad \min_{A \subset \mathcal{X}} \frac{\mathsf{Cut}(A)}{\mathsf{Vol}(A)\mathsf{Vol}(\mathcal{X} \setminus A)},$$

where

$$\mathsf{Vol}(A) = \sum_{x \in A} \sum_{y \in X} w_{xy}.$$

Gives good clusterings but very computationally hard (NP-hard).

Spectral clustering

For $A \subset \mathcal{X}$ set

$$u(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise.} \end{cases}$$

Spectral clustering

For $A \subset \mathcal{X}$ set

$$u(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise.} \end{cases}$$

Then we have

$$\text{Cut}(A) = \sum_{\substack{x,y\in\mathcal{X} \\ x\in A, y\notin A}} w_{xy} = \frac{1}{2}\sum_{x,y\in\mathcal{X}} w_{xy}(u(x)-u(y))^2$$

and

$$\text{Vol(A)} = \sum_{x,y\in\mathcal{X}} w_{xy}u(x).$$

Spectral clustering

For $A \subset \mathcal{X}$ set

$$u(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise.} \end{cases}$$

Then we have

$$\text{Cut}(A) = \sum_{\substack{x,y\in\mathcal{X} \\ x\in A, y\notin A}} w_{xy} = \frac{1}{2}\sum_{x,y\in\mathcal{X}} w_{xy}(u(x)-u(y))^2$$

and

$$\text{Vol(A)} = \sum_{x,y\in\mathcal{X}} w_{xy}u(x).$$

This allow us to write the balanced cut problem as

$$\min_{u:\mathcal{X}\to\{0,1\}} \frac{\sum_{x,y\in\mathcal{X}} w_{xy}(u(x)-u(y))^2}{\sum_{x,y,x',y'\in\mathcal{X}} u(x)w_{xy}(1-u(y'))w_{x'y'}}.$$

Spectral clustering

Consider solving the similar, relaxed, problem

$$\min_{\substack{u:\mathcal{X}\to\mathbb{R} \\ \sum_{x\in\mathcal{X}} u(x)\neq 0}} \frac{\displaystyle\sum_{x,y\in\mathcal{X}} w_{xy}(u(x)-u(y))^2}{\displaystyle\sum_{x\in\mathcal{X}} u(x)^2}.$$

Spectral clustering

Consider solving the similar, relaxed, problem

$$\min_{\substack{u:\mathcal{X}\to\mathbb{R} \\ \sum_{x\in\mathcal{X}} u(x)\neq 0}} \frac{\displaystyle\sum_{x,y\in\mathcal{X}} w_{xy}(u(x)-u(y))^2}{\displaystyle\sum_{x\in\mathcal{X}} u(x)^2}.$$

The solution is the smallest non-trivial eigenvector (Fiedler vector) of the graph Laplacian

$$\Delta u(x) = \sum_{y\in\mathcal{X}} w_{xy}(u(x)-u(y)).$$

Spectral clustering

Consider solving the similar, relaxed, problem

$$\min_{\substack{u:\mathcal{X}\to\mathbb{R} \\ \sum_{x\in\mathcal{X}} u(x)\neq 0}} \frac{\displaystyle\sum_{x,y\in\mathcal{X}} w_{xy}(u(x)-u(y))^2}{\displaystyle\sum_{x\in\mathcal{X}} u(x)^2}.$$

The solution is the smallest non-trivial eigenvector (Fiedler vector) of the graph Laplacian

$$\Delta u(x) = \sum_{y\in\mathcal{X}} w_{xy}(u(x)-u(y)).$$

Binary spectral clustering:

1. Compute Fiedler vector $u : \mathcal{X} \to \mathbb{R}$.
2. Set $A = \{x \in \mathcal{X} \,:\, u(x) > 0\}$.

Spectral clustering

Spectral clustering: To cluster into k groups:

1. Compute first k eigenvectors of the graph Laplacian Δ:

$$u_1, \ldots, u_k : \mathcal{X} \to \mathbb{R}.$$

Spectral clustering

Spectral clustering: To cluster into k groups:

1. Compute first k eigenvectors of the graph Laplacian Δ:

$$u_1, \ldots, u_k : \mathcal{X} \to \mathbb{R}.$$

2. Define the spectral embedding $\Psi : \mathcal{X} \to \mathbb{R}^k$ by

$$\Psi(x) = (u_1(x), u_2(x), \ldots, u_k(x)).$$

Spectral clustering

Spectral clustering: To cluster into k groups:

1. Compute first k eigenvectors of the graph Laplacian Δ:

$$u_1, \ldots, u_k : \mathcal{X} \to \mathbb{R}.$$

2. Define the spectral embedding $\Psi : \mathcal{X} \to \mathbb{R}^k$ by

$$\Psi(x) = (u_1(x), u_2(x), \ldots, u_k(x)).$$

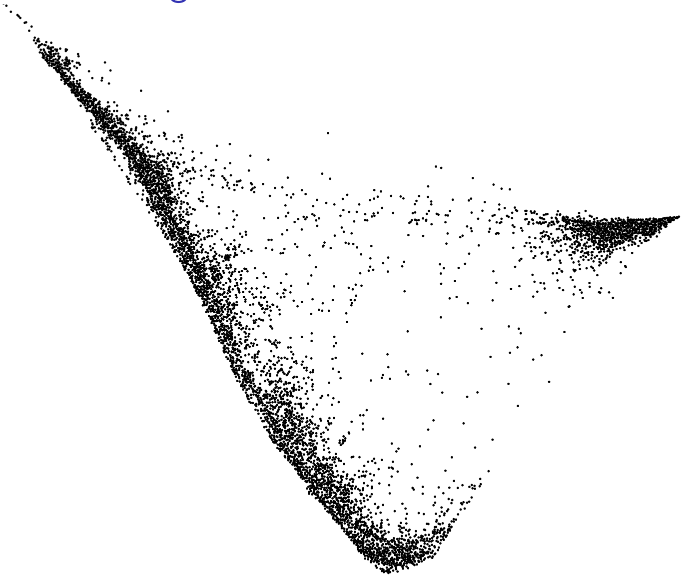
3. Cluster the point cloud $\mathcal{Y} = \Psi(\mathcal{X})$ with your favorite clustering algorithm (often k-means).

Spectral methods in data science

Spectral methods are widely used for dimension reduction and clustering in data science and machine learning.

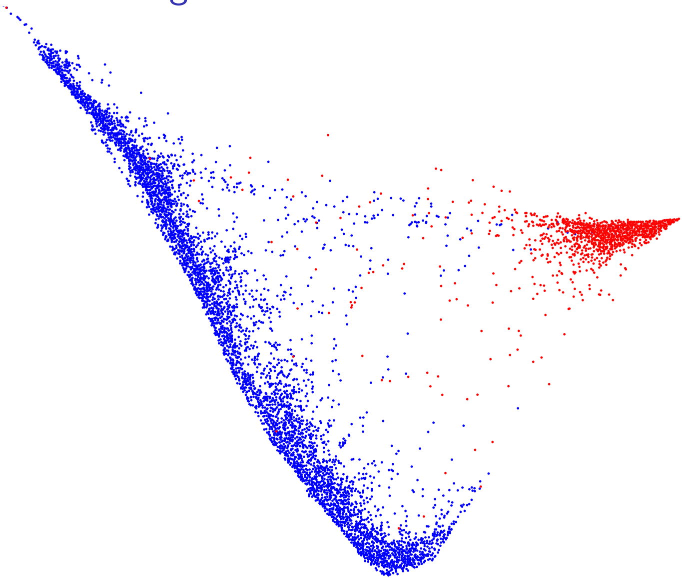
- Spectral clustering [Shi and Malik (2000)] [Ng, Jordan, and Weiss (2002)]
- Laplacian eigenmaps [Belkin and Niyogi (2003)]
- Diffusion maps [Coifman and Lafon (2006)]

Spectral embedding: MNIST



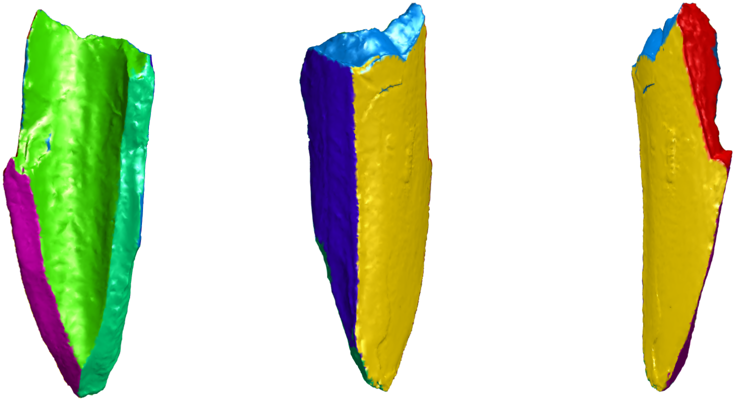
Digits 1 and 2 from MNIST visualized with spectral projection

Spectral embedding: MNIST



Digits 1 (blue) and 2 (red) from MNIST visualized with spectral projection

Application: Segmenting broken bone fragments



Spectral clustering with weights

$$w_{ij} = \exp\left(-C|\mathbf{n}_i - \mathbf{n}_j|^p\right).$$

between nearby points on the mesh, where $\mathbf{n}_i$ is the outward normal vector at vertex i.

Outline

Manifold assumption

Let $\mathcal{M} \subset \mathbb{R}^d$ be a compact, connected, orientable, smooth, m-dimensional manifold.

Manifold assumption

Let $\mathcal{M} \subset \mathbb{R}^d$ be a compact, connected, orientable, smooth, m-dimensional manifold.

We give to $\mathcal{M}$ the Riemannian structure induced by the ambient space $\mathbb{R}^d$. The geodesic distance between $x, y \in \mathcal{M}$ is denoted $d_{\mathcal{M}}(x,y)$ and

$$B_{\mathcal{M}}(x,r) = \{y \in \mathcal{M} \,:\, d_{\mathcal{M}}(x,y) < r\}.$$

By $dVol$ we denote the volume form on $\mathcal{M}$.

Manifold assumption

Let $\rho \in C^2(\mathcal{M})$, $\rho > 0$, and let

$$\mathcal{X}_n = \{x_1, \dots, x_n\}$$

be an i.i.d. sample from the distribution $\rho dVol_{\mathcal{M}}$.

Manifold assumption

Let $\rho \in C^2(\mathcal{M})$, $\rho > 0$, and let

$$\mathcal{X}_n = \{x_1, \ldots, x_n\}$$

be an i.i.d. sample from the distribution $\rho dVol_{\mathcal{M}}$.

Let $\eta\colon [0,\infty) \to [0,\infty)$ be non-increasing with

$$\eta(t) = 0 \quad \text{for} \quad t > 1.$$

We assume $\eta|_{[0,1]}$ is Lipschitz and that

$$\int_{\mathbb{R}^m} \eta(|w|)dw = 1,$$

Manifold assumption

Let $\rho \in C^2(\mathcal{M})$, $\rho > 0$, and let

$$\mathcal{X}_n = \{x_1, \ldots, x_n\}$$

be an i.i.d. sample from the distribution $\rho dVol_{\mathcal{M}}$.

Let $\eta\colon [0,\infty) \to [0,\infty)$ be non-increasing with

$$\eta(t) = 0 \quad \text{for} \quad t > 1.$$

We assume $\eta|_{[0,1]}$ is Lipschitz and that

$$\int_{\mathbb{R}^m} \eta(|w|)dw = 1,$$

Let $\varepsilon > 0$. The weights in the graph are

$$w_{xy} = \eta\left(\frac{|x-y|}{\varepsilon}\right).$$

The resulting graph is called a random geometric graph.

Spectral convergence

The spectrum of the graph-Laplacian converges ($n \to \infty, \varepsilon \to 0$) to the spectrum of the weighted Laplace-Beltrami operator

$$\Delta_{\mathcal{M}} u = -\rho^{-1} \mathrm{div}_{\mathcal{M}}(\rho^2 \nabla_{\mathcal{M}} u).$$

Spectral convergence

The spectrum of the graph-Laplacian converges ($n \to \infty, \varepsilon \to 0$) to the spectrum of the weighted Laplace-Beltrami operator

$$\Delta_{\mathcal{M}} u = -\rho^{-1}\mathrm{div}_{\mathcal{M}}(\rho^2 \nabla_{\mathcal{M}} u).$$

Spectral convergence results under manifold assumption:

- Belkin and Niyogi (2007)
- Shi (2015): $O(n^{-1/(4m+14)})$ rate in L^2.
- Trillos and Slepcev (2016)
- Singer and Wu (2017)
- Trillos, Gerlach, Hein, and Slepcev (2018): $O(n^{-1/4m})$ rate in L^2
- C., Trillos (2019): $O(n^{-1/(m+4)})$ rate in L^2
- Dunson, Wu, Wu (2019): $O(n^{-1/(4m+15)})$ rate in L^∞

Similar non-probabilistic results

- Fujiwara (1995), Burago, Ivanov and Kurylev (2014)

Outline of talk

Challenges for analysis:

- Spectral convergence results are hard because many useful PDE tools do not transfer to the graph-setting.

Outline of talk

Challenges for analysis:

- Spectral convergence results are hard because many useful PDE tools do not transfer to the graph-setting.
- Randomness in the graph can average out (homogenize) in ways that are difficult to analyze.

Outline of talk

Challenges for analysis:

- Spectral convergence results are hard because many useful PDE tools do not transfer to the graph-setting.
- Randomness in the graph can average out (homogenize) in ways that are difficult to analyze.

Question: What type of PDE tools (e.g., elliptic regularity) can we push to the random geometric graph setting?

Outline of talk

Challenges for analysis:

- Spectral convergence results are hard because many useful PDE tools do not transfer to the graph-setting.
- Randomness in the graph can average out (homogenize) in ways that are difficult to analyze.

Question: What type of PDE tools (e.g., elliptic regularity) can we push to the random geometric graph setting?

Today's talk: Lipschitz regularity for solutions of graph Poisson equations

$$\Delta u = f$$

and applications to spectral convergence.

Calder, J. and Garcia Trillos, N., Lewicka, M. **Lipschitz regularity of graph Laplacians on random data clouds**, *In preparation*, 2020.

Outline

Outline

Main results: Global Lipschitz regularity

Take the manifold assumption for $\mathcal{X}_n = \{x_1, x_2, \ldots, x_n\}$.

We define the graph Laplacian $\Delta_{\varepsilon,\mathcal{X}_n} : L^2(\mathcal{X}_n) \to L^2(\mathcal{X}_n)$ by

$$\Delta_{\varepsilon,\mathcal{X}_n} u(x_i) = \frac{1}{n\varepsilon^{m+2}} \sum_{j=1}^{n} \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right)\big(u(x_i) - u(x_j)\big).$$

Main results: Global Lipschitz regularity

Take the manifold assumption for $\mathcal{X}_n = \{x_1, x_2, \ldots, x_n\}$.

We define the graph Laplacian $\Delta_{\varepsilon,\mathcal{X}_n} : L^2(\mathcal{X}_n) \to L^2(\mathcal{X}_n)$ by

$$\Delta_{\varepsilon,\mathcal{X}_n} u(x_i) = \frac{1}{n\varepsilon^{m+2}} \sum_{j=1}^{n} \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right)(u(x_i) - u(x_j)).$$

Theorem (C., Garcia Trillos, Lewicka, 2020)

Let $\varepsilon \ll 1$. Then, with probability at least $1 - C\varepsilon^{-6m}\exp\left(-cn\varepsilon^{m+4}\right)$ we have

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

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

Main results: Interior Lipschitz regularity

We define the graph Laplacian $\Delta_{\varepsilon,\mathcal{X}_n} : L^2(\mathcal{X}_n) \to L^2(\mathcal{X}_n)$ by

$$\Delta_{\varepsilon,\mathcal{X}_n} u(x) = \frac{1}{n\varepsilon^{m+2}} \sum_{j=1}^{n} \eta\left(\frac{|x - x_j|}{\varepsilon}\right) (u(x) - u(x_j)).$$

Theorem (C., Garcia Trillos, Lewicka, 2020)

Let $0 < r < \mathrm{diam}(\mathcal{M})$ *where* $\mathrm{diam}(\mathcal{M})$ *is the diameter of* $\mathcal{M}$. *Then, for every* $\varepsilon > 0$ *satisfying* $\frac{(|\log(\varepsilon)|+1)\varepsilon}{r} \ll 1$, *with probability at least* $1 - C\varepsilon^{-6m}\exp\left(-cn\varepsilon^{m+4}\right)$ *we have*

$$|u(x_i) - u(x_j)| \leq C\|u\|_{L^\infty(\mathcal{X}_n \cap B_{\mathcal{M}}(x,7r))}\left(\varepsilon + \frac{|\log(\varepsilon)|\varepsilon}{r} + \frac{d_{\mathcal{M}}(x_i,x_j)}{r}\right) + C\varepsilon\|\Delta_{\varepsilon,\mathcal{X}_n} u\|_{L^\infty(\mathcal{X}_n \cap B_{\mathcal{M}}(x,7r))},$$

for all $u \in L^2(\mathcal{X}_n)$, $x \in \mathcal{M}$, $r > 0$, *and* $x_i, x_j \in B_{\mathcal{M}}(x,r) \cap \mathcal{X}_n$.

Main results: Lipschitz regularity of eigenvectors

Theorem (C., Garcia Trillos, Lewicka, 2020)

Let $\Lambda > 0$ *and* $\varepsilon \ll 1$, *and suppose that* $\varepsilon \leq \frac{c}{\Lambda+1}$. *Then, with probability at least* $1 - C\varepsilon^{-6m}\exp\left(-cn\varepsilon^{m+4}\right) - 2n\exp\left(-cn(\Lambda+1)^{-m}\right)$ *we have*

$$|u(x_i) - u(x_j)| \leq C(\Lambda+1)^{m+1}\|u\|_{L^1(\mathcal{X}_n)}(d_{\mathcal{M}}(x_i,x_j)+\varepsilon)$$

valid for all non-identically zero $u \in L^2(\mathcal{X}_n)$ *with* $\lambda_u \leq \Lambda$ *and all* $x_i, x_j \in \mathcal{X}_n$. *Here,*

$$\lambda_u = \frac{\|\Delta_{\varepsilon,\mathcal{X}_n}u\|_{L^\infty(\mathcal{X}_n)}}{\|u\|_{L^\infty(\mathcal{X}_n)}}.$$

Main results: Lipschitz regularity of eigenvectors

Theorem (C., Garcia Trillos, Lewicka, 2020)

Let $\Lambda > 0$ *and* $\varepsilon \ll 1$, *and suppose that* $\varepsilon \leq \frac{c}{\Lambda+1}$. *Then, with probability at least* $1 - C\varepsilon^{-6m}\exp\left(-cn\varepsilon^{m+4}\right) - 2n\exp\left(-cn(\Lambda+1)^{-m}\right)$ *we have*

$$|u(x_i) - u(x_j)| \leq C(\Lambda+1)^{m+1}\|u\|_{L^1(\mathcal{X}_n)}(d_{\mathcal{M}}(x_i, x_j) + \varepsilon)$$

valid for all non-identically zero $u \in L^2(\mathcal{X}_n)$ *with* $\lambda_u \leq \Lambda$ *and all* $x_i, x_j \in \mathcal{X}_n$. *Here,*

$$\lambda_u = \frac{\|\Delta_{\varepsilon,\mathcal{X}_n} u\|_{L^\infty(\mathcal{X}_n)}}{\|u\|_{L^\infty(\mathcal{X}_n)}}.$$

Corollary (C., Garcia Trillos, Lewicka, 2020)

Under the same conditions as above

$$\|u\|_{L^\infty(\mathcal{X}_n)} \leq C(\Lambda+1)^{m+1}\|u\|_{L^1(\mathcal{X}_n)},$$

for all u *non-identically zero with* $\lambda_u \leq \Lambda$.

Outline

Main results: Spectral convergence

Recall the continuum weighted Laplace-Beltrami operator.

$$\Delta_{\mathcal{M}} u(x) = -\rho^{-1} \mathrm{div}_{\mathcal{M}}(\rho^2 \nabla_{\mathcal{M}} u).$$

We also define

$$[u]_{\varepsilon,\mathcal{X}_n} = \max_{x,y \in \mathcal{X}_n} \frac{|u(x) - u(y)|}{d_{\mathcal{M}}(x,y) + \varepsilon}.$$

Theorem (C., Garcia Trillos, Lewicka, 2020)

Let $\varepsilon \ll 1$ and suppose that $u_{n,\varepsilon}$ is a normalized eigenvector of $\Delta_{\varepsilon,\mathcal{X}_n}$. Then, with probability at least $1 - C(n + \varepsilon^{-6m}) \exp\left(-cn\varepsilon^{m+4}\right)$ there exists a normalized eigenfunction u of $\Delta_{\mathcal{M}}$ for which

$$\|u_{n,\varepsilon} - u\|_{L^\infty(\mathcal{X}_n)} + [u_{n,\varepsilon} - u]_{\varepsilon,\mathcal{X}_n} \leq C\varepsilon,$$

where the constant C depends on u, $\mathcal{M}$, ρ.

Main results: Spectral convergence

Recall the continuum weighted Laplace-Beltrami operator.

$$\Delta_{\mathcal{M}} u(x) = -\rho^{-1}\mathrm{div}_{\mathcal{M}}(\rho^2 \nabla_{\mathcal{M}} u).$$

We also define

$$[u]_{\varepsilon,\mathcal{X}_n} = \max_{x,y\in\mathcal{X}_n} \frac{|u(x)-u(y)|}{d_{\mathcal{M}}(x,y)+\varepsilon}.$$

Theorem (C., Garcia Trillos, Lewicka, 2020)

Let $\varepsilon \ll 1$ and suppose that $u_{n,\varepsilon}$ is a normalized eigenvector of $\Delta_{\varepsilon,\mathcal{X}_n}$. Then, with probability at least $1 - C(n+\varepsilon^{-6m})\exp\left(-cn\varepsilon^{m+4}\right)$ there exists a normalized eigenfunction u of $\Delta_{\mathcal{M}}$ for which

$$\|u_{n,\varepsilon} - u\|_{L^\infty(\mathcal{X}_n)} + [u_{n,\varepsilon} - u]_{\varepsilon,\mathcal{X}_n} \le C\varepsilon,$$

where the constant C depends on u, $\mathcal{M}$, ρ.

Optimal choice for ε satisfies $n\varepsilon^{m+4} = C\log(n)$, which gives rates $O(n^{-1/(m+4)})$.

Outline

Outline

Outline of proof

Main ideas:

1. We lift the problem from the graph to the manifold $\mathcal{M}$ obtaining a related nonlocal Laplacian

$$\Delta_\varepsilon u(x) = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{d_{\mathcal{M}}(x,y)}{\varepsilon}\right) \big(u(x) - u(y)\big)\rho(y)\, dVol(y).$$

Outline of proof

Main ideas:

1. We lift the problem from the graph to the manifold $\mathcal{M}$ obtaining a related nonlocal Laplacian

$$\Delta_\varepsilon u(x) = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{d_{\mathcal{M}}(x,y)}{\varepsilon}\right)\big(u(x) - u(y)\big)\rho(y)\, dVol(y).$$

2. We prove the Lipschitz estimate for Δ_ε using a specific coupling of suitable random walks.
   - The coupling is based on the reflection coupling of [Lindvall & Rogers, 1986], with additional ingredients to handle a drift term.

Outline

Lifting to the manifold

Recall the graph Laplacian

$$\Delta_{\varepsilon,\mathcal{X}_n} u(x) = \frac{1}{n\varepsilon^{m+2}} \sum_{j=1}^{n} \eta\left(\frac{|x - x_j|}{\varepsilon}\right)(u(x) - u(x_j)).$$

If $u : \mathcal{M} \to \mathbb{R}$ is a smooth function, then we can compute for any $x \in \mathcal{M}$

$$\mathbb{E}[\Delta_{\varepsilon,\mathcal{X}_n} u(x)] = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{|x - y|}{\varepsilon}\right)(u(x) - u(y))\rho(y)\, dVol(y).$$

Lifting to the manifold

Recall the graph Laplacian

$$\Delta_{\varepsilon,\mathcal{X}_n} u(x) = \frac{1}{n\varepsilon^{m+2}} \sum_{j=1}^{n} \eta\left(\frac{|x - x_j|}{\varepsilon}\right) (u(x) - u(x_j)).$$

If $u : \mathcal{M} \to \mathbb{R}$ is a smooth function, then we can compute for any $x \in \mathcal{M}$

$$\mathbb{E}[\Delta_{\varepsilon,\mathcal{X}_n} u(x)] = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{|x - y|}{\varepsilon}\right) (u(x) - u(y))\rho(y)\, dVol(y).$$

An application of Bernstein's inequality yields

$$\mathbb{P}(|\Delta_{\varepsilon,\mathcal{X}_n} u(x) - \mathbb{E}[\Delta_{\varepsilon,\mathcal{X}_n} u(x)]| \geq C\text{Lip}(u)t) \leq 2\exp(-Cn\varepsilon^{m+2}t^2) \quad \text{for } 0 < t \leq 1$$

Theorem (Bernstein's inequality)

Let $Y_1, \ldots, Y_n$ be i.i.d. with mean $\mu = \mathbb{E}[Y_i]$ and variance $\sigma^2 = \mathbb{E}[(Y_i - \mathbb{E}[Y_i])^2]$, and assume $|Y_i| \leq M$ almost surely for all i. Then for any $t > 0$

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i - n\mu\right| > nt\right) \leq 2\exp\left(-\frac{nt^2}{2\sigma^2 + 4Mt/3}\right).$$

Lifting to the manifold

Recall the graph Laplacian

$$\Delta_{\varepsilon,\mathcal{X}_n} u(x_i) = \frac{1}{n\varepsilon^{m+2}} \sum_{j=1}^{n} \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right)(u(x_i) - u(x_j)).$$

If $u : \mathcal{M} \to \mathbb{R}$ is a smooth function, then we can compute for any $x \in \mathcal{M}$

$$\mathbb{E}[\Delta_{\varepsilon,\mathcal{X}_n} u(x)] = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{|x-y|}{\varepsilon}\right)(u(x) - u(y))\rho(y)\, dVol(y).$$

An application of Bernstein's inequality yields

$$\mathbb{P}(|\Delta_{\varepsilon,\mathcal{X}_n} u(x) - \mathbb{E}[\Delta_{\varepsilon,\mathcal{X}_n} u(x)]| \geq C\text{Lip}(u)t) \leq 2\exp(-Cn\varepsilon^{m+2}t^2) \quad \text{for } 0 < t \leq 1$$

For $\varepsilon \ll 1$ and $|x - y| \leq \varepsilon$ we have

$$|x - y| \leq d_{\mathcal{M}}(x, y) \leq |x - y| + O(\varepsilon^3).$$

Therefore

$$\mathbb{P}(|\Delta_{\varepsilon,\mathcal{X}_n} u(x) - \Delta_\varepsilon u(x)| \geq C\text{Lip}(u)t + C\varepsilon) \leq 2\exp(-Cn\varepsilon^{m+2}t^2) \quad \text{for } 0 < t \leq 1$$

Pointwise consistency

As an aside, for smooth functions u, the nonlocal Laplacian

$$\Delta_\varepsilon u(x) = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{d_{\mathcal{M}}(x,y)}{\varepsilon}\right)(u(x)-u(y))\rho(y)\, dVol(y)$$

is consistent with a weighted Laplace-Beltrami operator

$$\Delta_{\mathcal{M}} u(x) = -\rho^{-1}\mathrm{div}_{\mathcal{M}}(\rho^2 \nabla_{\mathcal{M}} u).$$

Pointwise consistency

As an aside, for smooth functions u, the nonlocal Laplacian

$$\Delta_\varepsilon u(x) = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{d_{\mathcal{M}}(x,y)}{\varepsilon}\right)(u(x) - u(y))\rho(y)\, dVol(y)$$

is consistent with a weighted Laplace-Beltrami operator

$$\Delta_{\mathcal{M}} u(x) = -\rho^{-1}\mathrm{div}_{\mathcal{M}}(\rho^2 \nabla_{\mathcal{M}} u).$$

Indeed, by Taylor expanding u we can show that

$$\Delta_\varepsilon u(x) = \sigma_\eta \Delta_{\mathcal{M}} u(x) + O(\varepsilon \|u\|_{C^3}).$$

Pointwise consistency

As an aside, for smooth functions u, the nonlocal Laplacian

$$\Delta_\varepsilon u(x) = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{d_{\mathcal{M}}(x,y)}{\varepsilon}\right)(u(x)-u(y))\rho(y)\, dVol(y)$$

is consistent with a weighted Laplace-Beltrami operator

$$\Delta_{\mathcal{M}} u(x) = -\rho^{-1}\mathrm{div}_{\mathcal{M}}(\rho^2 \nabla_{\mathcal{M}} u).$$

Indeed, by Taylor expanding u we can show that

$$\Delta_\varepsilon u(x) = \sigma_\eta \Delta_{\mathcal{M}} u(x) + O(\varepsilon \|u\|_{C^3}).$$

This gives pointwise consistency of graph Laplacians [Hein 2007]

$$\mathbb{P}(|\Delta_{\varepsilon,\mathcal{X}_n} u(x) - \sigma_\eta \Delta_{\mathcal{M}} u(x)| \geq C\mathrm{Lip}(u)t + C\varepsilon\|u\|_{C^3}) \leq 2\exp(-Cn\varepsilon^{m+2}t^2).$$

Note this requires $n\varepsilon^{m+2} \gg 1$, and for $t=\varepsilon$ we need $n\varepsilon^{m+4} \gg 1$.

Interpolation

We define the interpolation operator $\mathcal{I}_{\varepsilon,\mathcal{X}_n} : L^2(\mathcal{X}_n) \to L^2(\mathcal{M})$ and the degree

$$\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) = \frac{1}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{i=1}^{n} \eta\left(\frac{|x - x_i|}{\varepsilon}\right) u(x_i),$$

where $d_{\varepsilon,\mathcal{X}_n}(x)$ is the degree of x, given by

$$d_{\varepsilon,\mathcal{X}_n}(x) = \sum_{i=1}^{n} \eta\left(\frac{|x - x_i|}{\varepsilon}\right).$$

Interpolation

We define the interpolation operator $\mathcal{I}_{\varepsilon,\mathcal{X}_n} : L^2(\mathcal{X}_n) \to L^2(\mathcal{M})$ and the degree

$$\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) = \frac{1}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{i=1}^{n} \eta\left(\frac{|x - x_i|}{\varepsilon}\right) u(x_i),$$

where $d_{\varepsilon,\mathcal{X}_n}(x)$ is the degree of x, given by

$$d_{\varepsilon,\mathcal{X}_n}(x) = \sum_{i=1}^{n} \eta\left(\frac{|x - x_i|}{\varepsilon}\right).$$

Theorem (C., Garcia Trillos, Lewicka, 2020)

Let $\varepsilon \ll 1$. Then, with probability at least $1 - C\varepsilon^{-6m} \exp\big(-cn\varepsilon^{m+4}\big)$ we have

$$|\Delta_\varepsilon(\mathcal{I}_{\varepsilon,\mathcal{X}_n} u)(x)| \leq C\big(\|\Delta_{\varepsilon,\mathcal{X}_n} u\|_{L^\infty(\mathcal{X}_n \cap B(x,\varepsilon))} + \operatorname*{osc}_{\mathcal{X}_n \cap B(x,2\varepsilon)} u\big)$$

for all $u \in L^2(\mathcal{X}_n)$ and all $x \in \mathcal{M}$.

$$\Delta_\varepsilon u(x) = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{d_{\mathcal{M}}(x,y)}{\varepsilon}\right) \big(u(x) - u(y)\big)\rho(y)\, dVol(y).$$

Interpolation

We define the interpolation operator $\mathcal{I}_{\varepsilon,\mathcal{X}_n} : L^2(\mathcal{X}_n) \to L^2(\mathcal{M})$ and the degree

$$\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) = \frac{1}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{i=1}^{n} \eta\left(\frac{|x - x_i|}{\varepsilon}\right) u(x_i),$$

where $d_{\varepsilon,\mathcal{X}_n}(x)$ is the degree of x, given by

$$d_{\varepsilon,\mathcal{X}_n}(x) = \sum_{i=1}^{n} \eta\left(\frac{|x - x_i|}{\varepsilon}\right).$$

Corollary (C., Garcia Trillos, Lewicka, 2020)

Let $\varepsilon \ll 1$. With probability at least $1 - C\varepsilon^{-6m} \exp\left(-cn\varepsilon^{m+4}\right)$ we have

$$\|\Delta_\varepsilon(\mathcal{I}_{\varepsilon,\mathcal{X}_n} u)\|_{L^\infty(\mathcal{M})} \leq C\left(\|\Delta_{\varepsilon,\mathcal{X}_n} u\|_{L^\infty(\mathcal{X}_n)} + \varepsilon\|u\|_{L^\infty(\mathcal{X}_n)}\right)$$

for all $u \in L^2(\mathcal{X}_n)$.

$$\Delta_\varepsilon u(x) = \frac{1}{\varepsilon^{m+2}} \int_{\mathcal{M}} \eta\left(\frac{d_{\mathcal{M}}(x,y)}{\varepsilon}\right) (u(x) - u(y))\rho(y)\, dVol(y).$$

Interpolation: Proof sketch

Let $u : \mathcal{X}_n \to \mathbb{R}$ and denote $f(x) = u(x) - \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x)$ and $\eta_\varepsilon(t) = \varepsilon^{-d}\eta(t/\varepsilon)$. Then

$$\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) = \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^{n} \eta_\varepsilon(|x - x_j|)u(x_j)$$

Interpolation: Proof sketch

Let $u : \mathcal{X}_n \to \mathbb{R}$ and denote $f(x) = u(x) - \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x)$ and $\eta_\varepsilon(t) = \varepsilon^{-d}\eta(t/\varepsilon)$. Then

$$\begin{aligned}
\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) &= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|) u(x_j) \\
&= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|)\big(\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x_j) + f(x_j)\big)
\end{aligned}$$

Interpolation: Proof sketch

Let $u : \mathcal{X}_n \to \mathbb{R}$ and denote $f(x) = u(x) - \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x)$ and $\eta_\varepsilon(t) = \varepsilon^{-d}\eta(t/\varepsilon)$. Then

$$
\begin{aligned}
\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) &= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|) u(x_j) \\
&= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|)\big(\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x_j) + f(x_j)\big) \\
&= \varepsilon^{2d} \sum_{k=1}^n \left[\sum_{j=1}^n \frac{\eta_\varepsilon(|x - x_j|)\eta_\varepsilon(|x_j - x_k|)}{d_{\varepsilon,\mathcal{X}_n}(x) d_{\varepsilon,\mathcal{X}_n}(x_j)} \right] u(x_k) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x)
\end{aligned}
$$

Interpolation: Proof sketch

Let $u : \mathcal{X}_n \to \mathbb{R}$ and denote $f(x) = u(x) - \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x)$ and $\eta_\varepsilon(t) = \varepsilon^{-d}\eta(t/\varepsilon)$. Then

$$\begin{aligned}
\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) &= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|) u(x_j) \\
&= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|) \big(\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x_j) + f(x_j)\big) \\
&= \varepsilon^{2d} \sum_{k=1}^n \left[\sum_{j=1}^n \frac{\eta_\varepsilon(|x - x_j|)\eta_\varepsilon(|x_j - x_k|)}{d_{\varepsilon,\mathcal{X}_n}(x) d_{\varepsilon,\mathcal{X}_n}(x_j)} \right] u(x_k) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x) \\
&\approx \frac{1}{n\rho(x)} \sum_{k=1}^n \left[\int_{\mathcal{M}} \eta_\varepsilon(|x - y|)\eta_\varepsilon(|y - x_k|)\, d Vol_{\mathcal{M}}(y) \right] u(x_k) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x)
\end{aligned}$$

Interpolation: Proof sketch

Let $u : \mathcal{X}_n \to \mathbb{R}$ and denote $f(x) = u(x) - \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x)$ and $\eta_\varepsilon(t) = \varepsilon^{-d}\eta(t/\varepsilon)$. Then

$$\begin{aligned}
\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) &= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|) u(x_j) \\
&= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|)\big(\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x_j) + f(x_j)\big) \\
&= \varepsilon^{2d} \sum_{k=1}^n \left[\sum_{j=1}^n \frac{\eta_\varepsilon(|x - x_j|)\eta_\varepsilon(|x_j - x_k|)}{d_{\varepsilon,\mathcal{X}_n}(x) d_{\varepsilon,\mathcal{X}_n}(x_j)} \right] u(x_k) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x) \\
&\approx \frac{1}{n\rho(x)} \sum_{k=1}^n \left[\int_{\mathcal{M}} \eta_\varepsilon(|x - y|)\eta_\varepsilon(|y - x_k|)\, dVol_{\mathcal{M}}(y) \right] u(x_k) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x) \\
&= \frac{1}{n\rho(x)} \int_{\mathcal{M}} \eta_\varepsilon(|x - y|) \left[\sum_{k=1}^n \eta_\varepsilon(|y - x_k|) u(x_k) \right] dVol_{\mathcal{M}}(y) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x)
\end{aligned}$$

Interpolation: Proof sketch

Let $u : \mathcal{X}_n \to \mathbb{R}$ and denote $f(x) = u(x) - \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x)$ and $\eta_\varepsilon(t) = \varepsilon^{-d}\eta(t/\varepsilon)$. Then

$$
\begin{aligned}
\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) &= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|) u(x_j) \\
&= \frac{\varepsilon^d}{d_{\varepsilon,\mathcal{X}_n}(x)} \sum_{j=1}^n \eta_\varepsilon(|x - x_j|) \left(\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x_j) + f(x_j)\right) \\
&= \varepsilon^{2d} \sum_{k=1}^n \left[\sum_{j=1}^n \frac{\eta_\varepsilon(|x - x_j|)\eta_\varepsilon(|x_j - x_k|)}{d_{\varepsilon,\mathcal{X}_n}(x) d_{\varepsilon,\mathcal{X}_n}(x_j)} \right] u(x_k) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x) \\
&\approx \frac{1}{n\rho(x)} \sum_{k=1}^n \left[\int_{\mathcal{M}} \eta_\varepsilon(|x - y|)\eta_\varepsilon(|y - x_k|) \, dVol_{\mathcal{M}}(y) \right] u(x_k) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x) \\
&= \frac{1}{n\rho(x)} \int_{\mathcal{M}} \eta_\varepsilon(|x - y|) \left[\sum_{k=1}^n \eta_\varepsilon(|y - x_k|) u(x_k) \right] dVol_{\mathcal{M}}(y) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x) \\
&\approx \frac{1}{\rho(x)} \int_{\mathcal{M}} \eta_\varepsilon(|x - y|) \rho(y) \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(y) \, dVol_{\mathcal{M}}(y) + \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x)
\end{aligned}
$$

Interpolation: Proof sketch

Let $u : \mathcal{X}_n \to \mathbb{R}$ and denote $f(x) = u(x) - \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x)$ and $\eta_\varepsilon(t) = \varepsilon^{-d}\eta(t/\varepsilon)$. Then

$$\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) - \frac{1}{\rho(x)} \int_{\mathcal{M}} \eta_\varepsilon(|x-y|)\rho(y)\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(y) \, dVol_{\mathcal{M}}(y) \approx \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x).$$

Interpolation: Proof sketch

Let $u : \mathcal{X}_n \to \mathbb{R}$ and denote $f(x) = u(x) - \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x)$ and $\eta_\varepsilon(t) = \varepsilon^{-d}\eta(t/\varepsilon)$. Then

$$\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x) - \frac{1}{\rho(x)} \int_{\mathcal{M}} \eta_\varepsilon(|x-y|)\rho(y)\mathcal{I}_{\varepsilon,\mathcal{X}_n} u(y)\, dVol_{\mathcal{M}}(y) \approx \mathcal{I}_{\varepsilon,\mathcal{X}_n} f(x).$$

Then we check that

$$f(x_i) = u(x_i) - \mathcal{I}_{\varepsilon,\mathcal{X}_n} u(x_i) = \frac{n\varepsilon^{m+2}}{d_{\varepsilon,\mathcal{X}_n}(x_i)} \Delta_{\varepsilon,\mathcal{X}_n} u(x_i)$$

and

$$u(x) - \frac{1}{\rho(x)} \int_{\mathcal{M}} \eta_\varepsilon(|x-y|)\rho(y)u(y)\, dVol_{\mathcal{M}}(y) = \frac{\varepsilon^2}{\rho(x)} \Delta_\varepsilon u(x) + O(\varepsilon^2 \|u\|_\infty).$$

Interpolation

Theorem (C., Garcia Trillos, Lewicka, 2020)

Let $\varepsilon \ll 1$. Then, with probability at least $1 - C\varepsilon^{-6m} \exp\left(-cn\varepsilon^{m+4}\right)$ we have

$$|\Delta_\varepsilon(\mathcal{I}_{\varepsilon,\mathcal{X}_n} u)(x)| \leq C\left(\|\Delta_{\varepsilon,\mathcal{X}_n} u\|_{L^\infty(\mathcal{X}_n \cap B(x,\varepsilon))} + \operatorname*{osc}_{\mathcal{X}_n \cap B(x,2\varepsilon)} u\right)$$

for all $u \in L^2(\mathcal{X}_n)$ and all $x \in \mathcal{M}$.

Outline

Nonlocal operator

We have now lifted the problem to the manifold, and can assume u satisfies the mean-value type property

$$u(x) = \frac{1}{\rho(x)} \int_{B_{\mathcal{M}}(x,\varepsilon)} \eta_\varepsilon(|x-y|)\rho(y)u(y)\, dVol_{\mathcal{M}}(y) + \varepsilon^2 f(x)$$

for all $x \in \mathcal{M}$. The length scale $\varepsilon > 0$ is fixed and small.

Nonlocal operator

We have now lifted the problem to the manifold, and can assume u satisfies the mean-value type property

$$u(x) = \frac{1}{\rho(x)} \int_{B_{\mathcal{M}}(x,\varepsilon)} \eta_\varepsilon(|x-y|)\rho(y)u(y)\, dVol_{\mathcal{M}}(y) + \varepsilon^2 f(x)$$

for all $x \in \mathcal{M}$. The length scale $\varepsilon > 0$ is fixed and small.

We prove an approximate Lipschitz estimate for u depending on $\|u\|_\infty$ and $\|f\|_\infty$:

$$|u(x) - u(y)| \leq C(\|u\|_\infty + \|f\|_\infty)(d_{\mathcal{M}}(x,y) + \varepsilon).$$

- The proof uses the method of coupled random walks, similar to [Lindvall & Rogers, 1986].

Nonlocal operator

We have now lifted the problem to the manifold, and can assume u satisfies the mean-value type property

$$u(x) = \frac{1}{\rho(x)} \int_{B_{\mathcal{M}}(x,\varepsilon)} \eta_\varepsilon(|x-y|)\rho(y)u(y)\, dVol_{\mathcal{M}}(y) + \varepsilon^2 f(x)$$

for all $x \in \mathcal{M}$. The length scale $\varepsilon > 0$ is fixed and small.

We prove an approximate Lipschitz estimate for u depending on $\|u\|_\infty$ and $\|f\|_\infty$:

$$|u(x) - u(y)| \leq C(\|u\|_\infty + \|f\|_\infty)(d_{\mathcal{M}}(x,y) + \varepsilon).$$

- The proof uses the method of coupled random walks, similar to [Lindvall & Rogers, 1986].
- At a high level, this is equivalent to doubling the variables and using comparison to bound $u(x) - u(y) \leq \varphi(x,y)$ for a suitable supersolution φ.

Sketch of proof: Simple random walk

Assume u satisfies the mean-value property

$$u(x) = \fint_{B(x,\varepsilon)} u(y)\, dy$$

for fixed $\varepsilon > 0$ and all $B(x, \varepsilon)$.

Sketch of proof: Simple random walk

Assume u satisfies the mean-value property

$$u(x) = \fint_{B(x,\varepsilon)} u(y)\, dy$$

for fixed $\varepsilon > 0$ and all $B(x, \varepsilon)$. Let $x, y \in \mathbb{R}^d$ and assume we wish to estimate

$$|u(x) - u(y)| \leq C|x - y| + \dots$$

Sketch of proof: Simple random walk

Assume u satisfies the mean-value property

$$u(x) = \fint_{B(x,\varepsilon)} u(y)\, dy$$

for fixed $\varepsilon > 0$ and all $B(x,\varepsilon)$. Let $x, y \in \mathbb{R}^d$ and assume we wish to estimate

$$|u(x) - u(y)| \leq C|x-y| + \dots$$

WLOG assume $x = te_d$ and $y = -te_d$, $t \geq \varepsilon$.

Sketch of proof: Simple random walk

Assume u satisfies the mean-value property

$$u(x) = \fint_{B(x,\varepsilon)} u(y)\, dy$$

for fixed $\varepsilon > 0$ and all $B(x,\varepsilon)$. Let $x, y \in \mathbb{R}^d$ and assume we wish to estimate

$$|u(x) - u(y)| \leq C|x-y| + \dots$$

WLOG assume $x = te_d$ and $y = -te_d$, $t \geq \varepsilon$. Let X_k, Y_k be coupled simple random walks with $X_0 = x$, $Y_0 = y$ and

$$\begin{aligned} X_k &= X_{k-1} + \varepsilon U_k \\ Y_k &= Y_{k-1} + \varepsilon(U_k - 2(U_k \cdot e_d)e_d), \end{aligned}$$

where $U_1, U_2, \dots,$ are i.i.d. random variables uniformly distributed on $B(0,1)$.

Sketch of proof: Simple random walk

Assume u satisfies the mean-value property

$$u(x) = \fint_{B(x,\varepsilon)} u(y)\, dy$$

for fixed $\varepsilon > 0$ and all $B(x,\varepsilon)$. Let $x, y \in \mathbb{R}^d$ and assume we wish to estimate

$$|u(x) - u(y)| \leq C|x-y| + \dots$$

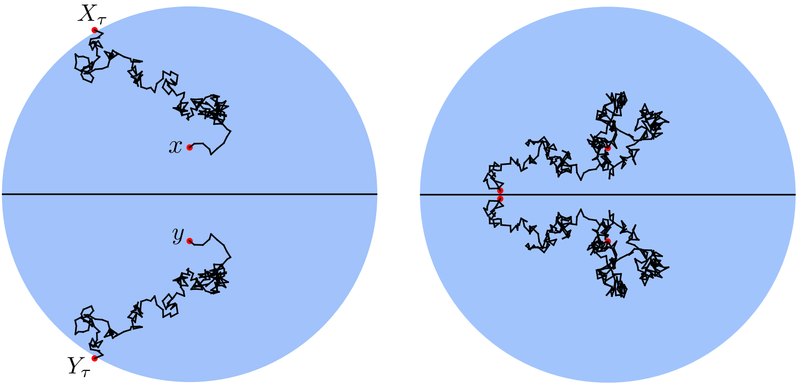
WLOG assume $x = te_d$ and $y = -te_d$, $t \geq \varepsilon$. Let X_k, Y_k be coupled simple random walks with $X_0 = x,\ Y_0 = y$ and

$$\begin{aligned} X_k &= X_{k-1} + \varepsilon U_k \\ Y_k &= Y_{k-1} + \varepsilon(U_k - 2(U_k \cdot e_d)e_d), \end{aligned}$$

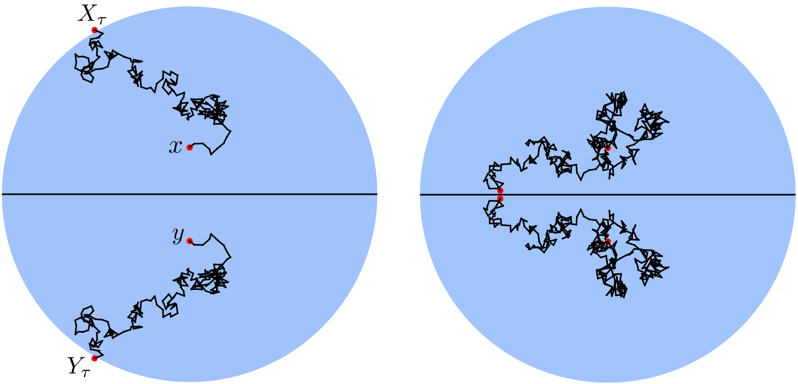
where $U_1, U_2, \dots,$ are i.i.d. random variables uniformly distributed on $B(0,1)$. For $r \gg t$, define the stopping time

$$\tau = \inf\left\{k > 0 \,:\, X_k \leq \frac{\varepsilon}{2} \text{ or } |X_k| > r\right\}.$$

Stopping time



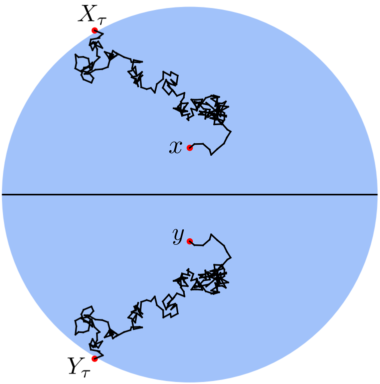
Stopping time



Since $u(X_k)$ and $u(Y_k)$ are martingales, Doob's optional stopping yields

$$u(x) - u(y) = \mathbb{E}[u(X_\tau) - u(Y_\tau)]$$

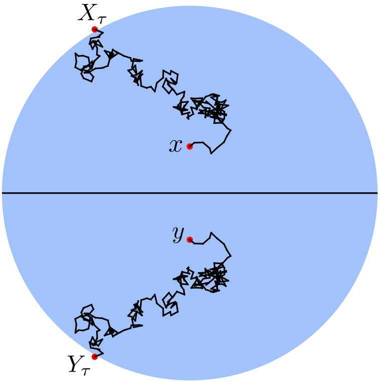
Exiting on $\partial B(0, r)$



If we have $|X_\tau| > r$ then we estimate

$$\mathbb{E}[u(X_\tau) - u(Y_\tau) \,|\, |X_\tau| > r] \leq 2\|u\|_{L^\infty(B(0,r+\varepsilon))}.$$

Exiting on $\partial B(0, r)$

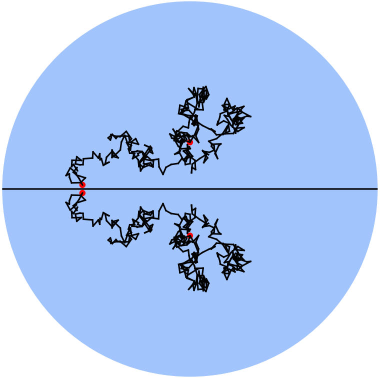


If we have $|X_\tau| > r$ then we estimate

$$\mathbb{E}[u(X_\tau) - u(Y_\tau)\,|\,|X_\tau| > r] \leq 2\|u\|_{L^\infty(B(0,r+\varepsilon))}.$$

$$\mathbb{P}(|X_\tau| > r) \leq \frac{Ct}{r} = C\frac{|x-y|}{r}.$$

Exiting on plane $x_d = 0$



If $|X_\tau| \leq r$, then $|X_\tau - Y_\tau| < \varepsilon$ and so

$$\mathbb{E}[u(X_\tau) - u(Y_\tau) \,|\, |X_\tau| \leq r] \leq \underbrace{\text{sup}\left\{|u(x') - u(y')| \,:\, x', y' \in B(0,r) \text{ and } |x' - y'| \leq \varepsilon\right\}}_{\Theta(r,\varepsilon)}.$$

Basic Lipschitz estimate

Conditioning on $|X_\tau| > r$ yields

$$\begin{aligned} u(x) - u(y) &= \mathbb{E}[u(X_\tau) - u(Y_\tau)] \\ &\leq 2\|u\|_{L^\infty(B(0,r+\varepsilon))}\mathbb{P}(|X_\tau| > r) + \Theta(r,\varepsilon)\mathbb{P}(|X_\tau| \leq r) \end{aligned}$$

Basic Lipschitz estimate

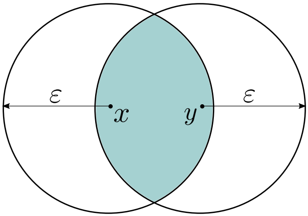
Conditioning on $|X_\tau| > r$ yields

$$\begin{aligned} u(x) - u(y) &= \mathbb{E}[u(X_\tau) - u(Y_\tau)] \\ &\leq 2\|u\|_{L^\infty(B(0,r+\varepsilon))}\mathbb{P}(|X_\tau| > r) + \Theta(r,\varepsilon)\mathbb{P}(|X_\tau| \leq r) \\ &\leq C\|u\|_{L^\infty(B(0,r+\varepsilon))}\frac{|x-y|}{r} + \Theta(r,\varepsilon). \end{aligned}$$

where

$$\Theta(r,\varepsilon) := \text{sup}\left\{|u(x') - u(y')| \,:\, x', y' \in B(0,r) \text{ and } |x'-y'| \leq \varepsilon\right\}.$$

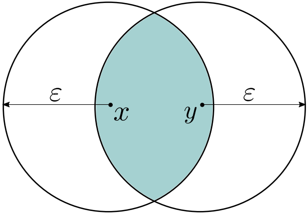
Local estimate



For x, y with $|x-y| \leq \varepsilon$ (and $x=-y$) we use the mean value property:

$$u(x) - u(y) \quad = \quad \frac{1}{|B(0,\varepsilon)|}\left(\int_{B(x,\varepsilon)} u(z)\,dz - \int_{B(y,\varepsilon)} u(z)\,dz\right)$$

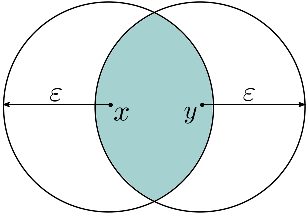
Local estimate



For x, y with $|x - y| \leq \varepsilon$ (and $x = -y$) we use the mean value property:

$$\begin{aligned} u(x) - u(y) &= \frac{1}{|B(0,\varepsilon)|}\left(\int_{B(x,\varepsilon)} u(z)\,dz - \int_{B(y,\varepsilon)} u(z)\,dz\right) \\ &= \frac{1}{|B(0,\varepsilon)|}\int_{B(x,\varepsilon)\setminus B(y,\varepsilon)} (u(x) - u(-x))\,dx \end{aligned}$$

Local estimate



For x, y with $|x-y| \leq \varepsilon$ (and $x = -y$) we use the mean value property:

$$\begin{aligned}
u(x) - u(y) &= \frac{1}{|B(0,\varepsilon)|}\left(\int_{B(x,\varepsilon)} u(z)\,dz - \int_{B(y,\varepsilon)} u(z)\,dz\right) \\
&= \frac{1}{|B(0,\varepsilon)|}\int_{B(x,\varepsilon)\setminus B(y,\varepsilon)} (u(x) - u(-x))\,dx \\
&\leq \eta \cdot \sup\left\{|u(x') - u(y')| \,:\, x', y' \in B(0, r+\varepsilon) \text{ and } |x'-y'| \leq 3\varepsilon\right\},
\end{aligned}$$

where $\eta = \dfrac{|B(x,\varepsilon) \setminus B(y,\varepsilon)|}{|B(0,\varepsilon)|} < 1$.

Global estimate

It follows that

$$\Theta(r,\varepsilon) := \text{sup}\left\{|u(x)-u(y)| \,:\, x,y\in B(0,r) \text{ and } |x-y|\leq\varepsilon\right\}$$

satisfies $\Theta(r,\varepsilon) \leq \eta\cdot\Theta(r+\varepsilon,3\varepsilon)$ for $\eta < 1$.

Global estimate

It follows that

$$\Theta(r,\varepsilon) := \text{sup}\left\{|u(x)-u(y)| \,:\, x,y\in B(0,r) \text{ and } |x-y|\leq\varepsilon\right\}$$

satisfies $\Theta(r,\varepsilon) \leq \eta\cdot\Theta(r+\varepsilon,3\varepsilon)$ for $\eta<1$. Thus, for $|x-y|\leq r$ we have

$$|u(x)-u(y)| \leq C\|u\|_{L^\infty(B(0,r+\varepsilon))}\frac{|x-y|}{r} + \eta\cdot\Theta(r+\varepsilon,3\varepsilon).$$

Global estimate

It follows that

$$\Theta(r,\varepsilon) := \sup\{|u(x)-u(y)| \,:\, x,y\in B(0,r) \text{ and } |x-y|\le\varepsilon\}$$

satisfies $\Theta(r,\varepsilon)\le \eta\cdot\Theta(r+\varepsilon,3\varepsilon)$ for $\eta<1$. Thus, for $|x-y|\le r$ we have

$$|u(x)-u(y)| \le C\|u\|_{L^\infty(B(0,r+\varepsilon))}\frac{|x-y|}{r} + \eta\cdot\Theta(r+\varepsilon,3\varepsilon).$$

On a periodic domain with no boundary (e.g., a closed manifold)

$$\Theta(r+\varepsilon,3\varepsilon)\le C\|u\|_{L^\infty}\varepsilon + \eta\cdot\Theta(r+\varepsilon,3\varepsilon),$$

and so

$$\Theta(r+\varepsilon,3\varepsilon)\le C(1-\eta)^{-1}\|u\|_{L^\infty}\varepsilon.$$

Global estimate

It follows that

$$\Theta(r,\varepsilon) := \text{sup}\left\{|u(x)-u(y)| \,:\, x,y\in B(0,r) \text{ and } |x-y|\leq\varepsilon\right\}$$

satisfies $\Theta(r,\varepsilon)\leq \eta\cdot\Theta(r+\varepsilon,3\varepsilon)$ for $\eta<1$. Thus, for $|x-y|\leq r$ we have

$$|u(x)-u(y)|\leq C\|u\|_{L^\infty(B(0,r+\varepsilon))}\frac{|x-y|}{r}+\eta\cdot\Theta(r+\varepsilon,3\varepsilon).$$

On a periodic domain with no boundary (e.g., a closed manifold)

$$\Theta(r+\varepsilon,3\varepsilon)\leq C\|u\|_{L^\infty}\varepsilon+\eta\cdot\Theta(r+\varepsilon,3\varepsilon),$$

and so

$$\Theta(r+\varepsilon,3\varepsilon)\leq C(1-\eta)^{-1}\|u\|_{L^\infty}\varepsilon.$$

This yields the global estimate

$$\boxed{|u(x)-u(y)|\leq C\|u\|_{L^\infty}(|x-y|+\varepsilon).}$$

Source terms

The argument extends directly to the inclusion of a source term

$$u(x) = \fint_{B(x,\varepsilon)} u(y)\, dy + \varepsilon^2 f(x).$$

Source terms

The argument extends directly to the inclusion of a source term

$$u(x) = \fint_{B(x,\varepsilon)} u(y)\, dy + \varepsilon^2 f(x).$$

In this case

$$Z_k = u(X_k) - u(Y_k) + \varepsilon^2 \|f\|_{L^\infty} k$$

is a submartingale, and Doob's optional stopping yields $Z_0 \leq \mathbb{E}[Z_\tau]$ or

$$u(x) - u(y) \leq \mathbb{E}[u(X_\tau) - u(Y_\tau)] + \varepsilon^2 \|f\|_{L^\infty} \mathbb{E}[\tau].$$

The proof proceeds similarly to obtain

$$|u(x) - u(y)| \leq C(\|u\|_{L^\infty} + \|f\|_{L^\infty})(|x - y| + \varepsilon).$$

Source terms

The argument extends directly to the inclusion of a source term

$$u(x) = \fint_{B(x,\varepsilon)} u(y)\, dy + \varepsilon^2 f(x).$$

In this case

$$Z_k = u(X_k) - u(Y_k) + \varepsilon^2 \|f\|_{L^\infty} k$$

is a submartingale, and Doob's optional stopping yields $Z_0 \leq \mathbb{E}[Z_\tau]$ or

$$u(x) - u(y) \leq \mathbb{E}[u(X_\tau) - u(Y_\tau)] + \varepsilon^2 \|f\|_{L^\infty} \mathbb{E}[\tau].$$

The proof proceeds similarly to obtain

$$|u(x) - u(y)| \leq C(\|u\|_{L^\infty} + \|f\|_{L^\infty})(|x-y| + \varepsilon).$$

Reference for simple random walk case: [Lewicka & Peres, 2019].

Coupled walks with drift

In the flat setting, our mean value property is

$$u(x) = \frac{1}{\rho(x)} \int_{B(x,\varepsilon)} \eta_\varepsilon(|x-y|)\rho(y)u(y)\,dy + \varepsilon^2 f(x).$$

Coupled walks with drift

In the flat setting, our mean value property is

$$u(x) = \frac{1}{\rho(x)} \int_{B(x,\varepsilon)} \eta_\varepsilon(|x-y|)\rho(y)u(y)\,dy + \varepsilon^2 f(x).$$

We Taylor expand $\rho(y) = \rho(x) + \nabla\rho(x)\cdot(y-x) + O(\varepsilon^2)$ to obtain

$$u(x) = \int_{B(x,\varepsilon)} \eta_\varepsilon(|x-y|)u(y)(1 + b(x)\cdot(y-x))\,dy + O(\varepsilon^2),$$

where $b(x) = \nabla \log \rho(x)$.

Coupled walks with drift

In the flat setting, our mean value property is

$$u(x) = \frac{1}{\rho(x)} \int_{B(x,\varepsilon)} \eta_\varepsilon(|x-y|)\rho(y)u(y)\,dy + \varepsilon^2 f(x).$$

We Taylor expand $\rho(y) = \rho(x) + \nabla\rho(x)\cdot(y-x) + O(\varepsilon^2)$ to obtain

$$u(x) = \int_{B(x,\varepsilon)} \eta_\varepsilon(|x-y|)u(y)(1 + b(x)\cdot(y-x))\,dy + O(\varepsilon^2),$$

where $b(x) = \nabla \log \rho(x)$. Assuming $\varepsilon|b(x)| \leq 1$ we can write

$$\begin{aligned} u(x) &= (1-\varepsilon|b(x)|)\int_{B(x,\varepsilon)} \eta_\varepsilon(|x-y|)u(y)\,dy \\ &\quad + \varepsilon|b(x)|\int_{B(x,\varepsilon)} \eta_\varepsilon(|x-y|)\left(1 - \frac{b(x)\cdot(y-x)}{\varepsilon|b(x)|}\right)u(y)\,dy + O(\varepsilon^2). \end{aligned}$$

Coupled walks with drift

Write $v(x) = \frac{b(x)}{|b(x)|}$ and $z = \frac{y-x}{\varepsilon}$ to simplify:

$$\begin{aligned} u(x) &= (1 - \varepsilon|b(x)|) \int_{B(0,1)} \eta(z) u(x + \varepsilon z)\, dz \\ &\quad + \varepsilon|b(x)| \int_{B(0,1)} \eta(z) \left(1 - v(x) \cdot z\right) u(x + \varepsilon z)\, dy + O(\varepsilon^2). \end{aligned}$$

Coupled walks with drift

Write $v(x) = \frac{b(x)}{|b(x)|}$ and $z = \frac{y-x}{\varepsilon}$ to simplify:

$$\begin{aligned} u(x) &= (1-\varepsilon|b(x)|)\int_{B(0,1)} \eta(z)u(x+\varepsilon z)\,dz \\ &\quad +\varepsilon|b(x)|\int_{B(0,1)} \eta(z)\,(1-v(x)\cdot z)\,u(x+\varepsilon z)\,dy + O(\varepsilon^2). \end{aligned}$$

Construction of coupled walks:

- Let $U_0, U_1, U_2, \ldots$ be i.i.d. with density $\eta(z)$.
- Let $V_0, V_1, V_2, \ldots$ be i.i.d. with density $\eta(z)(1-e_1\cdot z)$.
- Let $Q_0, Q_1, Q_2, \ldots$ be i.i.d. uniform on $[0,1]$.

Define $X_0 = x$ and

$$X_{k+1} = X_k + \varepsilon \begin{cases} U_k, & \text{if } Q_k > \varepsilon|b(X_k)| \\ O(e_1, v(X_k))\,V_k, & \text{otherwise,} \end{cases}$$

where $O(w,v)$ is an orthogonal matrix satisfying $O(w,v)w = v$.

Coupled walks with drift

Construction of coupled walks:

- Let $U_0, U_1, U_2, \ldots$ be i.i.d. with density $\eta(z)$.
- Let $V_0, V_1, V_2, \ldots$ be i.i.d. with density $\eta(z)(1 - e_1 \cdot z)$.
- Let $Q_0, Q_1, Q_2, \ldots$ be i.i.d. uniform on $[0, 1]$.

Define $X_0 = x$ and

$$X_{k+1} = X_k + \varepsilon \begin{cases} U_k, & \text{if } Q_k > \varepsilon|b(X_k)| \\ O(e_1, v(X_k))\, V_k, & \text{otherwise,} \end{cases}$$

where $O(w, v)$ is an orthogonal matrix satisfying $O(w, v)w = v$.

The coupled walk Y_k is constructed by setting $Y_0 = y$ and

$$Y_{k+1} = Y_k + \varepsilon \begin{cases} R(Y_k - X_k)\, U_k, & \text{if } Q_k > \varepsilon|b(Y_k)| \\ O(e_1, v(Y_k))\, V_k, & \text{otherwise,} \end{cases}$$

where $R(v)$ is a reflection matrix about the vector v.

Martingale property

Let $\mathcal{F}_k$ denote the σ-algebra induced by $U_0, \ldots, U_k$, $V_0, \ldots, V_k$, and $Q_0, \ldots, Q_k$. The coupled walks are constructed to have the approximate martingale property

$$\mathbb{E}[u(X_{k+1}) \,|\, \mathcal{F}_k] = \frac{1}{\rho(X_k)} \int_{B(X_k,\varepsilon)} \eta_\varepsilon(|X_k - y|)\rho(y)u(y)\, dy + O(\varepsilon^2).$$

$$\mathbb{E}[u(Y_{k+1}) \,|\, \mathcal{F}_k] = \frac{1}{\rho(Y_k)} \int_{B(Y_k,\varepsilon)} \eta_\varepsilon(|Y_k - y|)\rho(y)u(y)\, dy + O(\varepsilon^2).$$

Martingale property

Let $\mathcal{F}_k$ denote the σ-algebra induced by $U_0, \ldots, U_k$, $V_0, \ldots, V_k$, and $Q_0, \ldots, Q_k$. The coupled walks are constructed to have the approximate martingale property

$$\mathbb{E}[u(X_{k+1}) \,|\, \mathcal{F}_k] = \frac{1}{\rho(X_k)} \int_{B(X_k,\varepsilon)} \eta_\varepsilon(|X_k - y|)\rho(y)u(y)\,dy + O(\varepsilon^2).$$

$$\mathbb{E}[u(Y_{k+1}) \,|\, \mathcal{F}_k] = \frac{1}{\rho(Y_k)} \int_{B(Y_k,\varepsilon)} \eta_\varepsilon(|Y_k - y|)\rho(y)u(y)\,dy + O(\varepsilon^2).$$

The rest of the argument from the simple random walk setting goes through roughly the same.

Lifting to the manifold

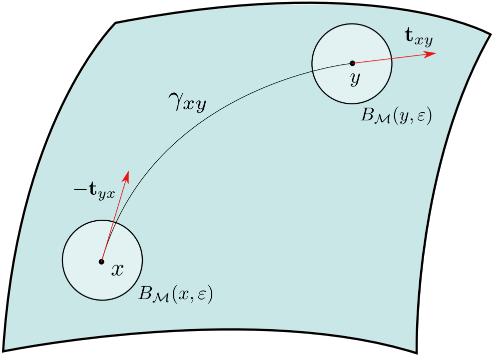
Our main result is in the (embedded) manifold setting $\mathcal{M} \subset \mathbb{R}^d$. In this case

$$u(x) = \frac{1}{\rho(x)} \int_{B_{\mathcal{M}}(x,\varepsilon)} \eta_\varepsilon(d_{\mathcal{M}}(x,y))\rho(y)u(y)\, dVol_{\mathcal{M}}(y) + \varepsilon^2 f(x).$$

Lifting to the manifold

Our main result is in the (embedded) manifold setting $\mathcal{M} \subset \mathbb{R}^d$. In this case

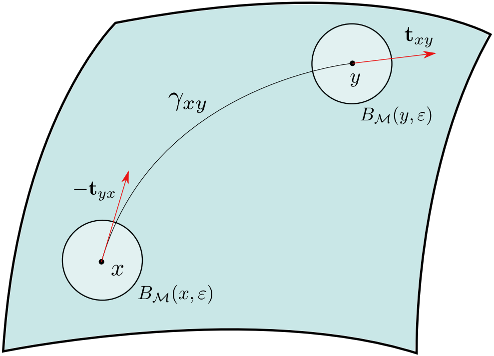
$$u(x) = \frac{1}{\rho(x)} \int_{B_{\mathcal{M}}(x,\varepsilon)} \eta_\varepsilon(d_{\mathcal{M}}(x,y))\rho(y)u(y)\, dVol_{\mathcal{M}}(y) + \varepsilon^2 f(x).$$



Lifting to the manifold

Our main result is in the (embedded) manifold setting $\mathcal{M} \subset \mathbb{R}^d$. In this case

$$u(x) = \frac{1}{\rho(x)} \int_{B_{\mathcal{M}}(x,\varepsilon)} \eta_\varepsilon(d_{\mathcal{M}}(x,y))\rho(y)u(y)\, dVol_{\mathcal{M}}(y) + \varepsilon^2 f(x).$$

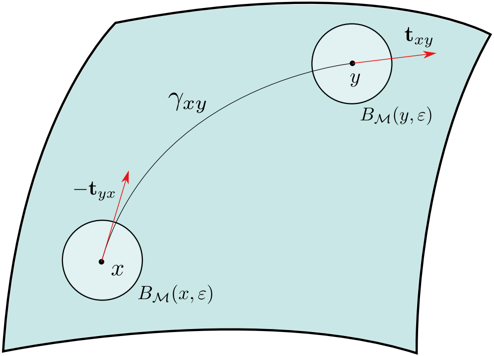


- γ_{xy} = geodesic from x to y.

Lifting to the manifold

Our main result is in the (embedded) manifold setting $\mathcal{M} \subset \mathbb{R}^d$. In this case

$$u(x) = \frac{1}{\rho(x)} \int_{B_{\mathcal{M}}(x,\varepsilon)} \eta_\varepsilon(d_{\mathcal{M}}(x,y))\rho(y)u(y)\, dVol_{\mathcal{M}}(y) + \varepsilon^2 f(x).$$



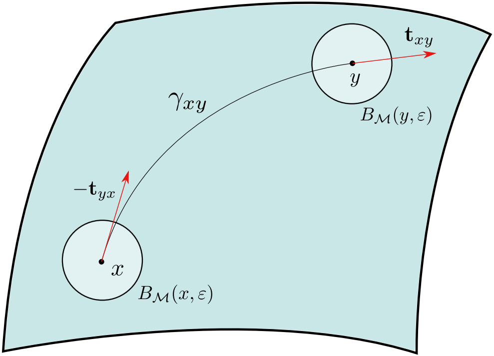
- γ_{xy} = geodesic from x to y.
- Define $\mathbf{t}_{xy} \in T_y\mathcal{M}$ by

$$\mathbf{t}_{xy} = \frac{d\gamma_{xy}}{ds}(d_{\mathcal{M}}(x,y)).$$

Lifting to the manifold

Our main result is in the (embedded) manifold setting $\mathcal{M} \subset \mathbb{R}^d$. In this case

$$u(x) = \frac{1}{\rho(x)} \int_{B_{\mathcal{M}}(x,\varepsilon)} \eta_\varepsilon(d_{\mathcal{M}}(x,y))\rho(y)u(y)\, dVol_{\mathcal{M}}(y) + \varepsilon^2 f(x).$$



- γ_{xy} = geodesic from x to y.
- Define $\mathbf{t}_{xy} \in T_y\mathcal{M}$ by

$$\mathbf{t}_{xy} = \frac{d\gamma_{xy}}{ds}(d_{\mathcal{M}}(x,y)).$$

- Let us denote by

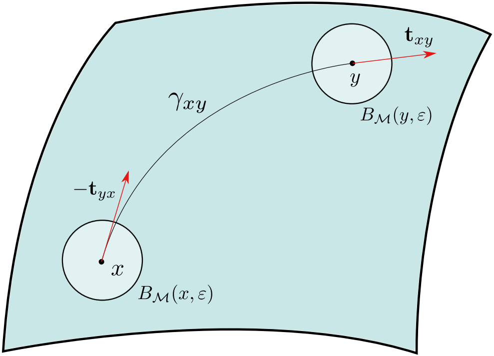
$$P_{xy} : T_x\mathcal{M} \to T_y\mathcal{M}$$

parallel transport along γ_{xy}.

Lifting to the manifold

Our main result is in the (embedded) manifold setting $\mathcal{M} \subset \mathbb{R}^d$. In this case

$$u(x) = \frac{1}{\rho(x)} \int_{B_{\mathcal{M}}(x,\varepsilon)} \eta_\varepsilon(d_{\mathcal{M}}(x,y))\rho(y)u(y)\, dVol_{\mathcal{M}}(y) + \varepsilon^2 f(x).$$



- γ_{xy} = geodesic from x to y.
- Define $\mathbf{t}_{xy} \in T_y\mathcal{M}$ by

$$\mathbf{t}_{xy} = \frac{d\gamma_{xy}}{ds}(d_{\mathcal{M}}(x,y)).$$

- Let us denote by

$$P_{xy} : T_x\mathcal{M} \to T_y\mathcal{M}$$

  parallel transport along γ_{xy}.
- Note that

$$\mathbf{t}_{xy} = P_{xy}(-\mathbf{t}_{yx}).$$

Coupled walks with drift on $\mathcal{M}$

Construction of coupled walks:

- Let $U_0, U_1, U_2, \ldots$ be i.i.d. with density $\eta(z)$.
- Let $V_0, V_1, V_2, \ldots$ be i.i.d. with density $\eta(z)(1 - e_1 \cdot z)$.
- Let $Q_0, Q_1, Q_2, \ldots$ be i.i.d. uniform on $[0, 1]$.

Define $X_0 = x$ and

$$X_{k+1} = \begin{cases} \exp_{X_k}(\varepsilon U_k), & \text{if } Q_k > \varepsilon |b(X_k)| \\ \exp_{X_k}(\varepsilon O(e_1, v(X_k)) V_k), & \text{otherwise,} \end{cases}$$

where $O(w, v)$ is an orthogonal matrix satisfying $O(w, v)w = v$.

The coupled walk Y_k is constructed by setting $Y_0 = y$ and

$$Y_{k+1} = \begin{cases} \exp_{Y_k}(\varepsilon R(\mathbf{t}_{X_k Y_k}) P_{X_k Y_k} U_k), & \text{if } Q_k > \varepsilon |b(Y_k)| \\ \exp_{Y_k}(\varepsilon O(e_1, v(Y_k)) P_{X_k Y_k} V_k), & \text{otherwise,} \end{cases}$$

where $R(v)$ is a reflection matrix about the vector v.

Outline

Future work

1. Similar estimates for other normalizations of the graph Laplacian
   - Random walk Laplacian

$$\Delta_{rw} u(x) = u(x) - \frac{1}{d_x} \sum_{y \in \mathcal{X}} w_{xy} u(y), \qquad d_x = \sum_{y \in \mathcal{X}} w_{xy}.$$

   - Normalized Laplacian

$$\Delta_{norm} u(x) = u(x) - \sum_{y \in \mathcal{X}} \frac{w_{xy}}{\sqrt{d_x d_y}} u(y).$$

Future work

1. Similar estimates for other normalizations of the graph Laplacian
   - Random walk Laplacian

$$\Delta_{rw} u(x) = u(x) - \frac{1}{d_x} \sum_{y \in \mathcal{X}} w_{xy} u(y), \qquad d_x = \sum_{y \in \mathcal{X}} w_{xy}.$$

   - Normalized Laplacian

$$\Delta_{norm} u(x) = u(x) - \sum_{y \in \mathcal{X}} \frac{w_{xy}}{\sqrt{d_x d_y}} u(y).$$

2. Other elliptic regularity results ($C^{1,\alpha}$, etc.).

Future work

1. Similar estimates for other normalizations of the graph Laplacian
   - Random walk Laplacian

$$\Delta_{rw} u(x) = u(x) - \frac{1}{d_x} \sum_{y \in \mathcal{X}} w_{xy} u(y), \qquad d_x = \sum_{y \in \mathcal{X}} w_{xy}.$$

   - Normalized Laplacian

$$\Delta_{norm} u(x) = u(x) - \sum_{y \in \mathcal{X}} \frac{w_{xy}}{\sqrt{d_x d_y}} u(y).$$

2. Other elliptic regularity results ($C^{1,\alpha}$, etc.).
3. Applications to other graph-based learning algorithms
   - Laplacian regularized semi-supervised learning.

Future work

1. Similar estimates for other normalizations of the graph Laplacian
   - Random walk Laplacian

$$\Delta_{rw} u(x) = u(x) - \frac{1}{d_x} \sum_{y \in \mathcal{X}} w_{xy} u(y), \qquad d_x = \sum_{y \in \mathcal{X}} w_{xy}.$$

   - Normalized Laplacian

$$\Delta_{norm} u(x) = u(x) - \sum_{y \in \mathcal{X}} \frac{w_{xy}}{\sqrt{d_x d_y}} u(y).$$

2. Other elliptic regularity results ($C^{1,\alpha}$, etc.).
3. Applications to other graph-based learning algorithms
   - Laplacian regularized semi-supervised learning.
4. Extending these results to smaller length scales using homogenization/percolation theory.

Outline

Length scale regimes

- Pointwise consistency of graph Laplacians requires

$$n\varepsilon^{m+2} \gg \log(n) \quad \iff \quad \varepsilon \gg \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+2}}.$$

Length scale regimes

- Pointwise consistency of graph Laplacians requires

$$n\varepsilon^{m+2} \gg \log(n) \quad \iff \quad \varepsilon \gg \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+2}}.$$

- Our Lipschitz regularity and $O(\varepsilon)$ spectral rates require

$$n\varepsilon^{m+4} \gg \log(n) \quad \iff \quad \varepsilon \gg \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+4}}.$$

Length scale regimes

- Pointwise consistency of graph Laplacians requires

$$n\varepsilon^{m+2} \gg \log(n) \quad \iff \quad \varepsilon \gg \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+2}}.$$

- Our Lipschitz regularity and $O(\varepsilon)$ spectral rates require

$$n\varepsilon^{m+4} \gg \log(n) \quad \iff \quad \varepsilon \gg \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+4}}.$$

- On the other hand, the graph is connected with high probability when

$$n\varepsilon^{m} \geq C\log(n) \quad \iff \quad \varepsilon \geq \left(\frac{C\log(n)}{n}\right)^{\frac{1}{m}}.$$

Some natural questions

Question 1: What can we say in the length scale regime

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

Some natural questions

Question 1: What can we say in the length scale regime

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

Question 2: What about smaller length scales

$$\varepsilon \sim \left(\frac{\log(n)}{n}\right)^{\frac{1}{m}}$$

where the graph is disconnected but has a giant component (supercritical percolation cluster)?

Some natural questions

Question 1: What can we say in the length scale regime

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

Question 2: What about smaller length scales

$$\varepsilon \sim \left(\frac{\log(n)}{n}\right)^{\frac{1}{m}}$$

where the graph is disconnected but has a giant component (supercritical percolation cluster)?

For Question 1, the Γ-convergence framework of Slepcev & Trillos establishes spectral convergence for $\varepsilon \gg \left(\frac{\log(n)}{n}\right)^{\frac{1}{m}}$, but the rates $O(\sqrt{\varepsilon})$ are far from sharp.

Some natural questions

Question 1: What can we say in the length scale regime

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

Question 2: What about smaller length scales

$$\varepsilon \sim \left(\frac{\log(n)}{n}\right)^{\frac{1}{m}}$$

where the graph is disconnected but has a giant component (supercritical percolation cluster)?

For Question 1, the Γ-convergence framework of Slepcev & Trillos establishes spectral convergence for $\varepsilon \gg \left(\frac{\log(n)}{n}\right)^{\frac{1}{m}}$, but the rates $O(\sqrt{\varepsilon})$ are far from sharp.

Do we expect, and can we prove, shaper rates?

Numerical experiments

Eigenmode	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Eigenvalue	2	2	2	6	6	6	6	6	12	12	12	12	12	12	12
E.value rate	2.4	2.6	3.1	2.3	2.3	2.5	2.6	3	2.1	2.1	2.2	2.3	2.4	2.8	3.3
E.vector rate	2.3	2.3	2.3	2.2	2.2	2.2	2.3	2.7	2.2	2.1	2.1	2.2	2.2	2.3	2.5

Table: Rates of convergence of the form $O(\varepsilon^b)$ (value of b is shown) for eigenvalues and eigenvectors of the graph Laplacian on the 2-sphere. Errors are averaged over 100 trials with n ranging from $n = 500$ to $n = 10^5$.

Rates of convergence for

$$\varepsilon = \left(\frac{\log n}{n}\right)^{\frac{1}{m+2}}.$$

At this length scale, our results give no convergence rate. For $O(\varepsilon^b)$ rate we require

$$\varepsilon \geq \left(\frac{\log n}{n}\right)^{\frac{1}{m+2+2b}}.$$

Homogenization at smaller length scales

The graph Laplacian

$$\Delta_{\varepsilon,\mathcal{X}_n} u(x) = \frac{1}{n\varepsilon^{m+2}} \sum_{j=1}^{n} \eta\left(\frac{|x - x_j|}{\varepsilon}\right)\left(u(x) - u(x_j)\right)$$

is not consistent with a continuum Laplacian when

$$\varepsilon \leq \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+2}}.$$

Homogenization at smaller length scales

The graph Laplacian

$$\Delta_{\varepsilon,\mathcal{X}_n} u(x) = \frac{1}{n\varepsilon^{m+2}} \sum_{j=1}^{n} \eta\left(\frac{|x - x_j|}{\varepsilon}\right)(u(x) - u(x_j))$$

is not consistent with a continuum Laplacian when

$$\varepsilon \leq \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+2}}.$$

However, we can construct other (homogenized) Laplacians that are consistent.

Homogenization at smaller length scales

Suppose $\Delta_{\varepsilon,\mathcal{X}_n} u \equiv 0$. Let $X_0, X_1, X_2, \ldots,$ be a random walk on the graph. Then $u(X_k)$ is a martingale and so for any k

$$u(x) = \mathbb{E}[u(X_k) \,|\, X_0 = x]$$

Homogenization at smaller length scales

Suppose $\Delta_{\varepsilon,\mathcal{X}_n} u \equiv 0$. Let $X_0, X_1, X_2, \ldots,$ be a random walk on the graph. Then $u(X_k)$ is a martingale and so for any k

$$u(x) = \mathbb{E}[u(X_k) \,|\, X_0 = x]$$

If we define

$$L_k u(x) := \mathbb{E}[u(x) - u(X_k) \,|\, X_0 = x].$$

Then $L_k u \equiv 0$.

Homogenization at smaller length scales

Suppose $\Delta_{\varepsilon,\mathcal{X}_n} u \equiv 0$. Let $X_0, X_1, X_2, \ldots,$ be a random walk on the graph. Then $u(X_k)$ is a martingale and so for any k

$$u(x) = \mathbb{E}[u(X_k) \,|\, X_0 = x]$$

If we define

$$L_k u(x) := \mathbb{E}[u(x) - u(X_k) \,|\, X_0 = x].$$

Then $L_k u \equiv 0$. L_k is a graph Laplacian; indeed, we can write

$$L_k u(x) = \sum_{i=1}^{n} \mathbb{P}(X_k = x_i \,|\, X_0 = x)(u(x) - u(x_i))$$

Homogenization at smaller length scales

Suppose $\Delta_{\varepsilon,\mathcal{X}_n} u \equiv 0$. Let $X_0, X_1, X_2, \ldots,$ be a random walk on the graph. Then $u(X_k)$ is a martingale and so for any k

$$u(x) = \mathbb{E}[u(X_k) \mid X_0 = x]$$

If we define

$$L_k u(x) := \mathbb{E}[u(x) - u(X_k) \mid X_0 = x].$$

Then $L_k u \equiv 0$. L_k is a graph Laplacian; indeed, we can write

$$L_k u(x) = \sum_{i=1}^{n} \mathbb{P}(X_k = x_i \mid X_0 = x)(u(x) - u(x_i))$$

The graph Laplacian L_k has effective length scale $\varepsilon_k = \varepsilon\sqrt{k}$. Hence, for $O(\varepsilon_k)$ pointwise consistency, we should only need

$$\varepsilon\sqrt{k} = \varepsilon_k \geq \left(\frac{\log(n)}{n}\right)^{\frac{1}{m+4}}.$$

Homogenization at smaller length scales

We can write this condition as

$$n\varepsilon^m(\varepsilon^4 k^{\frac{m+4}{m}}) \gg \log(n).$$

Homogenization at smaller length scales

We can write this condition as

$$n\varepsilon^m(\varepsilon^4 k^{\frac{m+4}{m}}) \gg \log(n).$$

If we assume $n\varepsilon^{m+p} \gg \log(n)$ for $p \geq 0$, then the smallest choice for k yields the effective length scale

$$\overline{\varepsilon} = \varepsilon_k = \varepsilon^{\frac{m+p}{m+4}}.$$

Homogenization at smaller length scales

We can write this condition as

$$n\varepsilon^m(\varepsilon^4 k^{\frac{m+4}{m}}) \gg \log(n).$$

If we assume $n\varepsilon^{m+p} \gg \log(n)$ for $p \geq 0$, then the smallest choice for k yields the effective length scale

$$\overline{\varepsilon} = \varepsilon_k = \varepsilon^{\frac{m+p}{m+4}}.$$

If we take $p = 0$, we are at graph connectivity, and

$$\overline{\varepsilon} = \varepsilon_k = \varepsilon^{\frac{m}{m+4}}.$$

So we expect nearly linear rates even at the smallest length scales.

Homogenization at smaller length scales

We can write this condition as

$$n\varepsilon^m(\varepsilon^4 k^{\frac{m+4}{m}}) \gg \log(n).$$

If we assume $n\varepsilon^{m+p} \gg \log(n)$ for $p \geq 0$, then the smallest choice for k yields the effective length scale

$$\overline{\varepsilon} = \varepsilon_k = \varepsilon^{\frac{m+p}{m+4}}.$$

If we take $p = 0$, we are at graph connectivity, and

$$\overline{\varepsilon} = \varepsilon_k = \varepsilon^{\frac{m}{m+4}}.$$

So we expect nearly linear rates even at the smallest length scales.

All of this requires proving Gaussian estimates on the heat kernel

$$p_k(x, x_i) = \mathbb{P}(X_k = x_i \mid X_0 = x)$$

when ε is small.