Classical, Graph and Nonlocal Laplacians

Rich Lehoucq
GraphEx2018
Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Adminstration under contract DE-NA0003525.

Outline

- Graph Laplacian; problems of interest
- Classical and nonlocal Laplacian, nonlocal vector calculus, volume-constrained problem the nonlocal analogue of a boundary-value problem
- Discrete vector calculus and problems of interest
- Graph nonlocality, edge expansion, expander families

Goal: demonstrate that continuum Laplacian ideas provide an interesting mathematical structure for graphs; perhaps a useful prism or guide?

Graph Laplacian

- Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a weighted, connected, undirected graph with n vertices
- $u \colon \ell^2(\mathcal{V}) \to \mathbb{R}, \quad \gamma \colon \ell^2(\mathcal{V} \times \mathcal{V}) \to \mathbb{R}^{0,+}$
- Define the graph (scaled) Laplacian $\mathcal{L}$

$$\mathcal{L}u(x) := -\frac{1}{\sigma(x)} \sum_y u(y)\, \gamma(x, y) + u(x)$$

$$\gamma(x, y) = \gamma(y, x), \quad \sigma(x) = \sum_y \gamma(x, y) > 0$$

- x, y are graph vertices, $\gamma(x, y)$ is the edge weight
- Can identify $\mathcal{V}$ with $\mathbb{R}^n$ so that $\mathcal{L}$ has a matrix representation $\mathbf{L} \in \mathbb{R}^{n \times n}$; $\mathbf{L} = \mathbf{D}^{-1}\mathbf{A} - \mathbf{I}$ for an unweighted graph

Classical and nonlocal Laplacian

- $u\colon L^2(\mathbb{R}^d) \to \mathbb{R}, \quad \gamma\colon L^2(\mathbb{R}^d \times \mathbb{R}^d) \to \mathbb{R}^{0,+}$ for a positive integer d; notation on u, γ and $\mathcal{L}$ is overloaded!
- Define the nonlocal Laplacian $\mathcal{L}$

$$\mathcal{L}u(x) := -\frac{1}{\sigma(x)} \int_{\mathbb{R}^d} u(y)\,\gamma(x,y)\,dy + u(x)$$

$$\gamma(x,y) = \gamma(y,x), \quad \sigma(x) = \int_{\mathbb{R}^d} \gamma(x,y)\,dy > 0$$

- $\mathcal{L}$ is a nonlocal operator because $y \neq x$ are needed in contrast to the classical Laplacian Δ given by

$$\Delta u(x) = \left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_d^2}\right) u(x), \quad x = (x_1, \ldots, x_d)$$

Relationship with the classical Laplacian

- The normalization $0 < \sigma(x) = \int_{\mathbb{R}} \gamma(x,y)\,dy < \infty$ rewrites

$$\begin{aligned} \mathcal{L}u(x) &= -\frac{1}{\sigma(x)} \int_{\mathbb{R}^d} u(y)\,\gamma(x,y)\,dy + u(x) \\ &= -\frac{1}{\sigma(x)} \int_{\mathbb{R}^d} \big(u(y) - u(x)\big)\,\gamma(x,y)\,dy \end{aligned}$$

- Selecting $\gamma(x,y) = \frac{\partial^2}{\partial y^2}\delta(y-x)$ where $\delta(y-x)$ is the Dirac delta measure grants that

$$\int_{\mathbb{R}^d} \big(u(y) - u(x)\big)\,\gamma(x,y)\,dy = \Delta u(x)$$

in a distributional sense

Motivation for the nonlocal operator

- Anomalous diffusion—Fick's first law doesn't hold
- Probabilistic—sample paths are discontinuous and so the infinitesimal generator contains a "jump" term
- Continuum analogue of the graph Laplacian
- Lovász and Szegedy *Limits of dense graph sequences*, J. Comb. Thy. 2006 implies that the limit is the nonlocal Laplacian

Nonlocal divergence operator

- Let $\boldsymbol{\alpha} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d, \boldsymbol{\alpha}(x, y) = -\boldsymbol{\alpha}(y, x)$ and $\mathbf{f} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$

$$\mathcal{B}\mathbf{f}(x) := \int_{\mathbb{R}^d} \big(\mathbf{f}(x, y) + \mathbf{f}(y, x)\big) \cdot \boldsymbol{\alpha}(x, y)\, dy$$

  where $\mathcal{B}\mathbf{f} : \mathbb{R}^d \to \mathbb{R}$

- $\mathcal{B}$ is a distributional divergence because

$$\boldsymbol{\alpha}(x, y) = -\frac{\partial}{\partial y}\delta(y - x) \Longrightarrow \mathcal{B}\mathbf{f}(x) \equiv \nabla \cdot \mathbf{f}(x, x)$$

  where $\nabla \cdot \mathbf{f}(x, x) := \frac{\partial}{\partial x_1}\mathbf{f}_1(x, x) + \cdots + \frac{\partial}{\partial x_d}\mathbf{f}_d(x, x)$

Adjoint of the nonlocal divergence operator

- Define the nonlocal gradient

$$\mathcal{B}^* u(x, y) := -\big(u(y) - u(x)\big)\, \alpha(x, y)$$

- Useful identity $\int_{\mathbb{R}^d} v \mathcal{B}\mathcal{B}^* u \, dx = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathcal{B}^* v \cdot \mathcal{B}^* u \, dy \, dx$
  Implication: $\mathcal{B}^*$ is the formal adjoint for $\mathcal{B}$
- Evocative of the classical relationship that the adjoint of $\nabla \cdot$ is $-\nabla$

Graph analogues of the divergence and gradient operators

- Set $\alpha(x,y) = \dfrac{1}{2\,\sigma(x)} \dfrac{x-y}{|x-y|} \sqrt{\gamma(x,y)}$

$$\mathcal{B}f(x) = \sum_y \big(f(x,y) + f(y,x)\big)\alpha(x,y)$$

$$\mathcal{B}^* u(x,y) = -\big(u(y) - u(x)\big)\,\alpha(x,y)$$

  where $\sigma(x) = \sum_y \gamma(x,y)$
- Can confirm that $\mathcal{B}\mathcal{B}^* u(x,y) = \mathcal{L}u(x)$
- Matrix representations of $\mathcal{B}, \mathcal{B}^*$ are the edge incidence matrix and its transpose

Application of nonlocal ideas to problems on Graphs

- See the notes *Analysis on Graphs* Grigoŕyan (Bielefeld), work over the last several years by Osher for applications to image processing and the recent paper *Statistical ranking and combinatorial Hodge theory*, Math. Prog., 2011.
- Although this is beautiful mathematics; is their a potential application? I believe this perspective is useful for us continuum solver folks to better understand graphs and assist in solution methods
- The relationship between the graph Laplacian and a Markov chain mimics the relationship between an elliptic PDE and Brownian motion

Hitting time random variable

- Hitting time of a vertex j is defined by the random variable

$$\mathcal{T}_{\partial\Omega} := \min\{j \geq 0 \colon \mathcal{X}_j \in \partial\Omega\},$$

  denoting the number of steps for the chain to visit $\partial\Omega = \{x_{i_1}, \ldots, x_{i_k}\}$
- Ω are the remaining vertices
- Mean hitting time starting at x is defined as

$$\mathbb{E}_x(\mathcal{T}_{\partial\Omega}) := \sum_{j=1}^{\infty} j\, \mathbb{P}_x(\mathcal{T}_{\partial\Omega} = j)$$

- Estimate by Monte Carlo on realizations of the Markov chain

Deterministic mean hitting time

- Classical consequence that the mean hitting time satisfies a symmetric positive definite set of linear equations

$$\begin{cases} \mathcal{L}\, h(x) = 1 & x \in \Omega \\ h(x) = 0 & x \in \partial\Omega \end{cases}$$

  where $h(x) = \mathbb{E}_x(\mathcal{T}_{\partial\Omega})$
- Analogous relationship that exists between Brownian motion and the solution of a Laplacian equation with homogenous Dirichlet boundary conditions
- And also the relationship between a jump diffusion and a volume constrained nonlocal Laplacian

Commute time τ_{yz}

- Let $\tau_{yz} = \mathbb{E}_y(\mathcal{T}_z) + \mathbb{E}_z(\mathcal{T}_y)$ is the expected number of steps to vertex z from y and then back to y
- Let $\left(u, v\right) := \sum_{x \in \mathcal{V}} u(x)v(x)\sigma(x)$. Then

$$\tau_{yz} = \min_{v \in \ell_0^2(\mathcal{V})} \frac{1}{2}\left(v, \mathcal{L}v\right) - \left(v, \delta_y - \delta_z\right)$$

  where $\delta_y(x) = 1$ for $x = y$ and zero otherwise; $\ell_0^2 := \{v \in \ell^2(\mathcal{V})\colon \sum_x v(x) = 0\}$
- The minimizer is given by the solution to the (singular) linear set of equations $\mathcal{L}u = \delta_y - \delta_z$
- Also have $\tau_{yz} = \left(u, \delta_y - \delta_z\right) = \left(\mathcal{L}^{-1}(\delta_y - \delta_z), \delta_y - \delta_z\right)$

Estimating the commute time

Suppose the graph is large enough so that we can only estimate τ_{yz}; therefore approximate

- **Method 1** Solution of the optimization problem
- **Method 2** Preconditioned iterative method the solution u of $\mathcal{L}u = \delta_y - \delta_z$ so that then $\tau_{yz} \approx \hat{u}(y) - \hat{u}(z)$
- **Method 3** the quadratic form $\left(\mathcal{L}^{-1}(\delta_y - \delta_z), \delta_y - \delta_z\right)$ in a number of eigenfunctions of $\mathcal{L}$
- **Method 4** using quadrature the expected value $\int \lambda^{-1} d\sigma_\lambda := \left(\mathcal{L}^{-1}(\delta_y - \delta_z), \delta_y - \delta_z\right)$ for a constructed measure σ_λ
- **Method 5** Compute the two mean hitting times (solve two symmetric positive definite linear systems)

Method 3

- Let $-\mathcal{L}\psi_i = \psi_i\lambda_i$ where $\lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{n-1} < 2$
- The commute time is equivalently expressed as

$$\tau_{yz} = \sum_{i=1}^{m} \frac{1}{\lambda_i} \mid \left(\delta_y - \delta_z, \psi_i\right) \mid^2 + \sum_{i=m+1}^{n-1} \frac{1}{\lambda_i} \mid \left(\delta_y - \delta_z, \psi_i\right) \mid^2$$

  where the hope is that $m \ll n$
- Important is the rate at which the eigenvalues λ_i^{-1} decay; slow decay implies m is in general not small

Preconditioning the classical Laplacian

- Discretizations of the boundary value problem

$$\begin{cases} \Delta u &= f \quad \text{on } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \end{cases}$$

  lead to systems of linear equations with coefficient matrix $\mathcal{L}$
- Preconditioning, in particular optimal preconditioning, is extremely well-understood
- Can we do this for a graph Laplacian?

Preconditioning

- Solving linear sets of equations with $\mathcal{L}$ and computing a number of eigenvalues and eigenfunctions of $\mathcal{L}$ is typically accomplished via an iterative method that approximates the solution or eigenpairs
- An iterative method is often accelerated via the use of a preconditioner $\hat{\mathcal{L}} \approx \mathcal{L}$; this can be made precise

$$c_1(v, \hat{\mathcal{L}} v) \leq (v, \mathcal{L} v) \leq c_2(v, \hat{\mathcal{L}} v) \quad \forall v \in \ell_0^2(\mathcal{V})$$

  for positive constants c_1 and c_2; $\hat{\mathcal{L}}$ is deemed an optimal preconditioner if c_1, c_2 are independent of n, the number of vertices
- The hope is that $c_1, c_2 \approx 1$ and that solving linear systems with $\hat{\mathcal{L}}$ is faster than with $\mathcal{L}$

Classifying graph generative processes

- Optimal preconditioning assumes that there is a generative process, e.g., discretization, for generating a sequence graph Laplacians; recall that a preconditioner $\hat{\mathcal{L}} \approx \mathcal{L}$; this can be made precise

$$c_1(v, \hat{\mathcal{L}}v) \leq (v, \mathcal{L}v) \leq c_2(v, \hat{\mathcal{L}}v) \quad \forall v \in \ell_0^2$$

  for positive constants c_1 and c_2; $\hat{\mathcal{L}}$ is deemed an optimal preconditioner if c_1, c_2 are independent of n, the number of vertices
- Can we classify a graph generative process so that we can give meaning to optimal preconditioning?
- This apparently can be done when $\mathcal{L}$ is associated with a complete graph because then the sequence of r-regular expanders do the job

Why are the reformulations useful?

- Discrete vector calculus allows a classification of the graph into flows, triangles etc. and go beyond the underlying pairwise assumption; the paper *Statistical ranking and combinatorial Hodge theory*, Math. Prog., 2011.
- Can consider exit-time and other random variables on graphs by the application of constraints
- The nonlocal Laplacian theory explains that the kernel γ tells you much about the nonlocal operator; consequently the graph kernel γ can explain the decay rate of the eigenvalues λ_i^{-1} useful for explaining how many eigenfunctions are needed
- For graph Laplacian's associated with a generative process, we can give meaning to optimal preconditioning

Relating a graph and nonlocality

- The operator $\int u(y)\,\gamma(x,y)\,dy + u(x)$ is nonlocal because $y \neq x$ are needed
- What is a graph consequence of nonlocality (outside of a jump diffusion process)
- Do the graphs associated with numerical discretizations of the nonlocal Laplacian possess special structure over those associated with the classical Laplacian?

Edge expansion

- The edge expansion (also isoperimetric number or Cheeger constant) is

$$h(\mathcal{G}) := \min_{0<|\mathcal{S}|<\frac{n}{2}} \frac{|\partial\mathcal{S}|}{|\mathcal{S}|}$$

- $\partial\mathcal{S}$ is the edge boundary of $\mathcal{S}$, i.e., the set of edges with exactly one endpoint in $\mathcal{S}$
- Graph analogue of the “surface” to volume ratio
- The “surface” is a bit of a misnomer—the edge boundary contains the edges connecting vertices within a subgraph and its complement

Expander families

- A family $\{\mathcal{G}_1, \mathcal{G}_2, \ldots\}$ of r-regular (unweighted) graphs is an expander family if there is a constant $c > 0$ such that $h(\mathcal{G}_i) \geq c$ for each $\mathcal{G}_i$
- Roughly, the “surface” to volume ratio of a graph does not go to zero as the graph increases in volume
- Roughly, families of graphs induced by discretizations of the classical Laplacian are not expander families because the “surface” to volume ratio goes to zero (as the graph increases in volume)
- In contrast, families of graphs induced by discretizations of the nonlocal Laplacian appear to be expander families because the “surface” to volume ratio does not go to zero (as the graph increases in volume)

Summary

- Graph Laplacian; problems of interest
- Classical and nonlocal Laplacian, nonlocal vector calculus, volume-constrained problem the nonlocal analogue of a boundary-value problem
- Discrete vector calculus and problems of interest
- Graph nonlocality, edge expansion, expander families

Goal: demonstrate that continuum Laplacian ideas can impact problems in graph analysis