Hamilton-Jacobi equations on graphs with applications to semi-supervised learning and data depth

Jeff Calder
School of Mathematics
University of Minnesota

The geometry of PDEs on graphs: analysis and applications
SIAM Conference on Analysis of PDEs
March 15, 2022

Joint work with Mahmood Ettehad (IMA)

Research supported by NSF grant DMS:1944925, the Alfred P. Sloan foundation, and a McKnight Presidential Fellowship.
Graph distance functions

Suppose we have a graph $G$ on $n$ vertices $\mathcal{X}$ with edge weights $w_{ij}$.

Set $I_n = \{1, \ldots, n\}$. The graph distance $d_G : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is defined by

$$d_G(x_i, x_j) = \min_{m \geq 1} \min_{\tau \in I_n^m} \left\{ \frac{w_{i, \tau_1}}{m} + \sum_{i=1}^{m-1} \frac{w_{\tau_i, \tau_{i+1}}}{m} + \frac{w_{\tau_m, j}}{m} \right\},$$
Graph distance functions: density weighting

The weighted graph distance $d_{G,f} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is defined by

\begin{equation}
(2) \quad d_{G,f}(x_i, x_j) := \min_{m \geq 1} \min_{\tau \in I^m_{n}} \left\{ w_{i,\tau_1}^{-1} f(x_{\tau_1}) + \sum_{i=1}^{m-1} w_{\tau_i,\tau_{i+1}}^{-1} f(x_{\tau_{i+1}}) + w_{\tau_m,\tau_j}^{-1} f(x_{\tau_j}) \right\}.
\end{equation}

It is common to choose $f = \hat{\rho}^{-\alpha}$, for some density estimation $\hat{\rho}$. 

Expensive
Prior work/References

Applications of graph distances:

- Dimensionality reduction (e.g., ISOMAP) [Tenenbaum et al., 2000]
- Semi-supervised learning on graphs, e.g., [Bijral, et al, 2003] [Chapelle and Zien, 2005]
- Graph classification [Borgwardt and Kriegel, 2005]
- Data depth [Calder, Park and Slepcev, 2021] [Molina-Fructuoso and Murray, 2022]

Discrete to continuum:

- $k$-nn graphs [Alamgir and Von Luxburg, 2012]
- Geodesic manifold distance [Hwang, Damelin, and Hero, 2016]
- Geodesic distance on Euclidean domains [Bungert, Calder, and Roith, 2021]
Lack of robustness to corrupted edges

Figure: From left to right we added an increasing number of corrupted edges (0, 10, 50, and 200) with edge weight $w_{ij} = 1$. 

(a) Graph distance function with corrupted edges
Recall

\begin{equation}
    d_{G,f}(x_i, x_j) := \min_{m \geq 1} \min_{\tau \in I_m^n} \left\{ w_{i,\tau_1}^{-1} f(x_{\tau_1}) + \sum_{i=1}^{m-1} w_{\tau_i,\tau_{i+1}}^{-1} f(x_{\tau_{i+1}}) + w_{\tau_m,j}^{-1} f(x_{\tau_j}) \right\}.
\end{equation}

Let us define the graph distance to a set $\Gamma$ by

$$
d_{G,f}(x, \Gamma) := \min_{x_j \in \Gamma} d_{G,f}(x_i, x_j).
$$

If $G$ is connected then $u(x) = d_{G,f}(x, \Gamma)$ is the unique solution of the graph eikonal equation

\begin{equation}
\begin{cases}
\max_{x_j \in X} w_{ji} (u(x_i) - u(x_j)) = f(x_i), & \text{if } x_i \in X \setminus \Gamma \\
\quad u(x_i) = 0, & \text{if } x_i \in \Gamma.
\end{cases}
\end{equation}
The p-eikonal equation

For $p > 0$, we define the $p$-eikonal operator $\mathcal{A}_{G,p} : F(\mathcal{X}) \to F(\mathcal{X})$ by

(5) \quad \mathcal{A}_{G,p} u(x_i) = \sum_{j=1}^{n} w_{ji} (u(x_i) - u(x_j))^p_+,

where $a_+ := \max\{a, 0\}$ is the positive part. For $\Gamma \subset \mathcal{X}$ and $f \in F(\mathcal{X})$, we consider the $p$-eikonal equation

(6) \quad \begin{cases} \mathcal{A}_{G,p} u = f, & \text{in } \mathcal{X} \setminus \Gamma \\ u = 0, & \text{on } \Gamma. \end{cases}

References:

- The $p$-eikonal equation originally appeared in [Desquenes, Elmoataz and Lezoray, 2013] with applications to image processing ($p = 1, 2, \infty$).
Let $K$ be the maximum unweighted degree of the graph, and $G^\alpha$ be the graph with weights $w^\alpha_{ij}$.

**Theorem (Well-posedness)**

Let $p > 0$ and $f > 0$. If $G$ is connected, then the $p$-eikonal equation

$$
\begin{cases}
A_{G,p} u = f, & \text{in } \mathcal{X} \setminus \Gamma \\
 u = 0, & \text{on } \Gamma.
\end{cases}
$$

has a unique solution $u \in F(\mathcal{X})$, and

$$
K^{-\frac{1}{p}} \left( \min_{\mathcal{X}} f^{\frac{1}{p}} \right) d_{G^{\frac{1}{p}}} (x_i, \Gamma) \leq u(x_i) \leq \left( \max_{\mathcal{X}} f^{\frac{1}{p}} \right) d_{G^{\frac{1}{p}}} (x_i, \Gamma).
$$

Note that the estimates above imply that $u$ recovers the graph distance as $p \to \infty$. 
Robustness

(a) Graph distance function with corrupted edges

(b) \( p \)-eikonal equation with \( p = 1 \) with corrupted edges
Robustness

Theorem (Robustness)

Let $\delta W$ have nonnegative entries, and set $\tilde{G} = (\mathcal{X}, W + \delta W)$ and $\delta G = (\mathcal{X}, \delta W)$. Let $u, \tilde{u} \in F(\mathcal{X})$ satisfy

\[
\begin{cases}
A_{\tilde{G}, p} \tilde{u}(x_i) = A_{G, p} u(x_i) = f(x_i), & \text{if } x_i \in \mathcal{X} \setminus \Gamma \\
\tilde{u}(x_i) = u(x_i) = 0, & \text{if } x_i \in \Gamma.
\end{cases}
\]

Then for all $x_i \in \mathcal{X}$ we have

\[
0 \leq \frac{u(x_i) - \tilde{u}(x_i)}{u(x_i)} \leq \left( \max_{\mathcal{X} \setminus \Gamma} \frac{A_{\delta G, p} u}{f} \right)^{\frac{1}{p}}.
\]

- The theorem can be simplified to give the weaker bound

\[
0 \leq \frac{u(x_i) - \tilde{u}(x_i)}{u(x_i)} \leq C \left( \frac{f_{\max}}{f_{\min}} \right)^{\frac{1}{p}} \| \delta W \|_{u,1}^{\frac{1}{p}},
\]

where

\[
\| A \|_{u,1} = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |A_{ij}| 1_{u(x_j) > u(x_i)}.
\]
Discrete to continuum

- Let $x_1, x_2, \ldots, x_n$ be a sequence of i.i.d random variables on $\Omega \subset \mathbb{R}^d$ with Lipschitz and positive density $\rho$ and set

  \[(11) \quad \mathcal{X} := \{x_1, x_2, \ldots, x_n\} \]

- Assume that $\Omega \subset \mathbb{R}^d$ is open, bounded and connected with a $C^{1,1}$ boundary $\partial \Omega$.

- We define the $p$-eikonal operator on a random geometric graph as

  \[\mathcal{A}_{n,\epsilon} u(x) := \frac{1}{n \sigma_p \epsilon^p} \sum_{y \in \mathcal{X}} \eta_{\epsilon}(|x - y|) (u(x) - u(y))^p_+ , \]

  where $\eta_{\epsilon}(t) := \frac{1}{\epsilon^d} \eta(\frac{t}{\epsilon})$ and set $\sigma_p := \int_{\mathbb{R}^d} \eta_{\epsilon}(|z|) |z_1|^p \, dz$

- Let $\Gamma \subset \mathcal{X}$ such that

  \[\text{dist}(\Gamma, \partial \Omega) \geq R,\]

  where $R$ is the reach of $\partial \Omega$. 
Discrete to continuum

For $p \geq 1$ we consider the $p$-eikonal equation with arbitrary right hand side $f$:

\[
\begin{cases}
A_{n,\varepsilon} u(x) = f(x) & \text{if } x \in \mathcal{X} \setminus \Gamma \\
u(x) = 0 & \text{if } x \in \Gamma.
\end{cases}
\]

**Continuum limit**: State-constrained eikonal equation [Capuzzo-Dolcetta & Lions, 1990]

\[
\begin{cases}
\rho |\nabla u|^p = f & \text{in } \Omega \setminus \Gamma \\
u = 0 & \text{on } \Gamma.
\end{cases}
\]
Define the geodesic weighted distance

\[ d_f(x, y) := \inf \left\{ \int_0^1 f(\gamma(t))|\gamma'(t)| \, dt : \gamma \in C^1([0, 1]; \Omega), \gamma(0) = x, \text{ and } \gamma(1) = y \right\}. \]

and set

\[ u(x) = \min_{y \in \Gamma} d_f(x, y). \]
Define the geodesic weighted distance

\[ d_f(x, y) := \inf \left\{ \int_0^1 f(\gamma(t))|\gamma'(t)|\,dt : \gamma \in C^1([0, 1]; \overline{\Omega}), \gamma(0) = x, \text{ and } \gamma(1) = y \right\}. \]

and set

\[ u(x) = \min_{y \in \Gamma} d_f(x, y). \]

Then \( u \) is the unique viscosity solution of the state constrained eikonal equation

\[ \begin{cases} 
|\nabla u| = f & \text{in } \Omega \setminus \Gamma \\
u = 0 & \text{on } \Gamma. 
\end{cases} \]

In particular, the solution of the continuum problem

\[ \begin{cases} 
\rho|\nabla u|^p = f & \text{in } \Omega \setminus \Gamma \\
u = 0 & \text{on } \Gamma.
\end{cases} \]

is given by \( u(x) = d_g(x, \Gamma) \), where \( g = \rho^{-\frac{1}{p}} f^{\frac{1}{p}} \).
Let $u_{n,\varepsilon}$ be the solution of

$$
\begin{cases}
A_{n,\varepsilon} u_{n,\varepsilon}(x) = f(x) & \text{if } x \in \mathcal{X} \setminus \Gamma \\
0 & \text{if } x \in \Gamma.
\end{cases}
$$

**Theorem**

There exists $C, c > 0$ such that for $\varepsilon$ sufficiently small and any $0 < \lambda \leq 1$ we have

$$
\mathbb{P} \left( \max_{x \in \mathcal{X}} (d_g(x, \Gamma) - u_{n,\varepsilon}(x)) \leq C(\sqrt{\varepsilon} + \lambda) \right) \geq 1 - 2n \exp(-cn\varepsilon^d \lambda^2).
$$

**Theorem**

There exists $C, c > 0$ such that for $\varepsilon$ sufficiently small and any $0 < \lambda \leq 1$ we have

$$
\mathbb{P} \left( \max_{x \in \mathcal{X}} (u_{n,\varepsilon}(x) - d_g(x, \Gamma)) \leq C \left( \sqrt{\varepsilon} + \left(n \varepsilon^{p+d} \right)^{\frac{1}{p}} + \lambda \right) \right) \geq 1 - 3n^2 \exp(-cn\varepsilon^d \lambda^2).
$$
Discrete to continuum

**Theorem**

There exists $C, c > 0$ such that for $\varepsilon$ sufficiently small and any $0 < \lambda \leq 1$ we have

$$
\mathbb{P} \left( \max_{x \in \mathcal{X}} (d_g(x, \Gamma) - u_{n,\varepsilon}(x)) \leq C(\sqrt{\varepsilon} + \lambda) \right) \geq 1 - 2n \exp(-cn\varepsilon^d \lambda^2).
$$

**Theorem**

There exists $C, c > 0$ such that for $\varepsilon$ sufficiently small and any $0 < \lambda \leq 1$ we have

$$
\mathbb{P} \left( \max_{x \in \mathcal{X}} (u_{n,\varepsilon}(x) - d_g(x, \Gamma)) \leq C \left( \sqrt{\varepsilon} + \left( n\varepsilon^{p+d} \right)^{\frac{1}{p}} + \lambda \right) \right) \geq 1 - 3n^2 \exp(-cn\varepsilon^d \lambda^2).
$$

If we choose $\lambda = \sqrt{\varepsilon}$ then we obtain that the convergence rate

$$
\max_{x \in \mathcal{X}} |u_{n,\varepsilon}(x) - d_g(x, \Gamma)| \leq C \left( \sqrt{\varepsilon} + \left( n\varepsilon^{p+d} \right)^{\frac{1}{p}} \right)
$$

holds with probability at least $1 - 5n^2 \exp(-cn\varepsilon^{d+1})$. 
Discrete to continuum

In order for the second result to be non-vacuous, we require that

\[ n\epsilon^d \gg \log(n) \quad \text{and} \quad n\epsilon^{d+p} \ll 1 \]

which can be reformulated as

\[ \left( \frac{\log(n)}{n} \right)^{\frac{1}{d}} \ll \epsilon \ll \left( \frac{1}{n} \right)^{\frac{1}{p+d}}. \]

For any \( p > 0 \) we can find feasible \( \epsilon \) (we use \( p \geq 1 \)).

A similar restriction appears in \( p \)-Laplacian learning with very few labels [Slepcev & Thorpe, 2019]

\[ \left( \frac{\log(n)}{n} \right)^{\frac{1}{d}} \ll \epsilon \ll \left( \frac{1}{n} \right)^{\frac{1}{p}}. \]

In this case, \( p > d \) is required to find feasible \( \epsilon \).
Discrete to continuum

(c) $\varepsilon = 0.03, p = 1$

(d) $\varepsilon = 0.06, p = 1$

(e) $\varepsilon = 0.09, p = 1$

(f) $\varepsilon = 0.03, p = 2$

(g) $\varepsilon = 0.06, p = 2$

(h) $\varepsilon = 0.09, p = 2$
Discrete to continuum

Main ideas in proof:

- Pointwise consistency $A_{n,\varepsilon}\varphi(x) \approx \rho|\nabla \varphi|^p$ for smooth $\varphi$, with high probability.

- The $O(\sqrt{\varepsilon})$ rate comes from a doubling variables argument in the viscosity solutions framework.

- Rate requires Lipschitzness of $u_{n,\varepsilon}$, we show that

$$|u_{n,\varepsilon}(x) - u_{n,\varepsilon}(y)| \leq c_p \gamma_p^{-1} \max_{\mathcal{X}} f_{p'} \frac{1}{p'} d_{\Omega}(x, y) + \gamma_p \left( n\varepsilon^{p+d} \right)^{1/p}, \text{ for all } x, y \in \mathcal{X}$$

with probability at least $1 - n^2 \exp \left( -\frac{c_d r_d^d}{2^{d+3}} \rho_{\min} n \varepsilon^d \right)$. The proof uses a geodesic cone barrier function with an additional spike:

$$v_{\beta, y}(x) := \beta(1 - \delta_y(x)) + d_{\Omega}(x, y)$$

- State constrained boundary condition handled with domain perturbation results.
Given a set $\Gamma \subset \mathcal{X}$ and a density estimation $\hat{\rho} : \mathcal{X} \to \mathbb{R}$, we consider solving the density reweighted $p$-eikonal equation

\begin{equation}
\label{eq:14}
\begin{cases}
A_{G,p} u = \hat{\rho}^{-\alpha}, & \text{in } \mathcal{X} \setminus \Gamma \\
u = 0, & \text{on } \Gamma,
\end{cases}
\end{equation}

where the exponent $\alpha$ is a tunable parameter. We denote the solution of \eqref{eq:14} by $D^{p,\alpha}_{\Gamma}(x) = u(x)$.

When $\Gamma = \{x\}$ is a single point we write $D^{p,\alpha}_{x}$. 
Data depth

Recall the **geometric median**:

\[ x_\ast \in \arg \min_{x \in \mathbb{R}^d} \sum_{i=1}^{n} |x_i - x|. \]

We can generalize this to the \( p \)-eikonal graph setting as follows:

\[ x_{p,\alpha} \in \arg \min_{x \in \mathcal{X}} \sum_{x_i \in \mathcal{X}} D_{x,\alpha}^{p,\alpha}(x_i). \]

Then we can define data depth as the distance to the median

\[ \text{depth}_{p,\alpha}(x) = \max_{x \in \mathcal{X}} D_{x_{p,\alpha}}^{p,\alpha} - D_{x_{p,\alpha}}^{p,\alpha}(x). \]

**Note:** Other approaches include first finding the “boundary” nodes and defining depth as distance to the boundary.

- [Calder, Park, & Slepcev, 2021]
- [Molina-Fructuoso and Murray, 2022]
Data depth

Figure: The $p$-eikonal medians and depth on 2D toy datasets with $p = 1$. The medians are shown for $\alpha = -1 \ (\triangledown)$, $\alpha = 0 \ (\Box)$ and the $\alpha = 1 \ (\triangle)$, while the points are colored by the $\alpha = 1$ data depth.
Data depth

(a) Helix  (b) Half Sphere  (c) Swiss Roll

**Figure:** The $p$-eikonal data depth on 3D toy datasets sampled from manifolds embedded in $\mathbb{R}^3$. We use $p = 1$ and $\alpha = 1$. 
Data depth

(a) Deepest images (median)

(b) Shallowest images (outliers)

Figure: Comparison of deepest (median) images to shallowest (outlier) images from each MNIST digit.
Data depth

(a) Deepest images (median)

(b) Shallowest images (outliers)

Figure: Comparison of deepest (median) images to shallowest (outlier) images from each FashionMNIST class.
Data depth

(a) MNIST
(b) FashionMNIST

Figure: Paths from shallowest point to median for each class.
Semi-supervised learning

- Suppose we have $k$ classes, and for each class $j = 1, \ldots, k$, we are provided some labeled nodes $\Gamma_j \subset \mathcal{X}$.

- The label prediction $\ell_i$ for an unlabeled node $x_i \not\in \Gamma_j$ for any $j$, is the label of the closest labeled node, under the distance $D_{\Gamma_j}^{p,\alpha}$, that is

$$\ell_i = \arg \min_{1 \leq j \leq k} D_{\Gamma_j}^{p,\alpha}(x_i).$$

- We can incorporate prior information about class sizes by introducing weights $s_j$ in the label decision [Calder et al, 2020]

$$\ell_i = \arg \min_{1 \leq j \leq k} \left\{ s_j D_{\Gamma_j}^{p,\alpha}(x_i) \right\}.$$
Figure: Comparison of the $p$-eikonal equation with $p = 1$ for semi-supervised image classification to Poisson learning [Calder et al., 2020] and the eikonal equation.
Semi-supervised learning

![Graph](image)

(a) CIFAR-10

(b) Accuracy vs $\alpha$

**Figure:** (a) Accuracy results for the $p$-eikonal equation with $p = 1$ for semi-supervised image classification on CIFAR-10, and (b) change in accuracy as the density reweighting exponent $\alpha$ is adjusted.
Paper and Code

Paper:


Code for all experiments is on GitHub

https://github.com/jwcalder/peikonal

The $p$-eikonal equation is implemented in the GraphLearning python package

https://github.com/jwcalder/GraphLearning  (pip install graphlearning)