Journal of Machine Learning Research 23 (2022) 1-62 Submitted 3/22; Revised 9/22; Published 10/22

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

Jeff Calder JCALDER@UMN.EDU
School of Mathematics
University of Minnesota
Mineapolis, MN 55455, USA

Mahmood Ettehad ETTEH001@UMN.EDU
Institute for Mathematics and its Applications (IMA)
University of Minnesota
Mineapolis, MN 55455, USA

Editor: Pradeep Ravikumar

Abstract

Shortest path graph distances are widely used in data science and machine learning, since they can approximate the underlying geodesic distance on the data manifold. However, the shortest path distance is highly sensitive to the addition of corrupted edges in the graph, either through noise or an adversarial perturbation. In this paper we study a family of Hamilton-Jacobi equations on graphs that we call the p-eikonal equation. We show that the p-eikonal equation with $p = 1$ is a provably robust distance-type function on a graph, and the $p \to \infty$ limit recovers shortest path distances. While the p-eikonal equation does not correspond to a shortest-path graph distance, we nonetheless show that the continuum limit of the p-eikonal equation on a random geometric graph recovers a geodesic density weighted distance in the continuum. We consider applications of the p-eikonal equation to data depth and semi-supervised learning, and use the continuum limit to prove asymptotic consistency results for both applications. Finally, we show the results of experiments with data depth and semi-supervised learning on real image datasets, including MNIST, FashionMNIST and CIFAR-10, which show that the p-eikonal equation offers significantly better results compared to shortest path distances.

Keywords: Data depth, Graph learning, Hamilton-Jacobi equation, Robust statistics, Semi-supervised learning, viscosity solutions, discrete to continuum limits, partial differential equations

1. Introduction

Shortest path distances on graphs have found applications in many areas of data science and machine learning, including dimensionality reduction (e.g., the ISOMAP algorithm (Tenenbaum et al., 2000)), semi-supervised learning on graphs (Moscovich et al., 2016; Chapelle and Zien, 2005; Bijral et al., 2012; Rozza et al., 2014; Yang et al., 2021), graph classification (Borgwardt and Kriegel, 2005), and data depth (Molina-Fructuoso and Murray, 2021, 2022; Calder et al., 2022b). In many applications, the shortest paths are density weighted, to make path lengths shorter in high density regions of the graph, and longer in

*. **Source Code:** `https://github.com/jwcalder/peikonal`

sparse regions (Bijral et al., 2012; Little et al., 2022). Shortest path algorithms offer different information compared to second order methods based on graph Laplacians, like spectral clustering (Ng et al., 2002), Laplacian eigenmaps (Belkin and Niyogi, 2003), diffusion maps (Coifman and Lafon, 2006), or Laplacian based semi-supervised learning (Zhu et al., 2003; Calder et al., 2020a), which offer information about average or typical paths through graphs. Some recent works have even combined shortest path metrics with graph Laplacian spectras to improve spectral clustering (Little et al., 2020).

However, a main drawback of shortest path distances is their lack of robustness to perturbations in graph structure. The addition of a single edge can have a strong effect on the shortest path, while simultaneously having little or no effect on the average or typical path, which gives an intuitive reason for the apparent superiority of graph Laplacian based methods for semi-supervised learning and dimension reduction, among other problems.

In this paper, we approach the problem of robustly computing distance functions on graphs from the viewpoint of Hamilton-Jacobi equations. We study a family of Hamilton-Jacobi equations on graphs, which we call the p-eikonal equations, that are provably robust to graph perturbations, especially for $p = 1$. The equations have the form

$$\sum_{j=1}^{n} w_{ji}(u(x_i) - u(x_j))_+^p = f(x_i), \tag{1}$$

where $a_+ = \max\{a, 0\}$, and w_{ij} is the weight between nodes i and j in the graph. We prove that as $p \to \infty$, these p-eikonal equations recover shortest path graph distances, while for $p = 1$ the solutions provide information that is different from shortest paths and far more robust to graph perturbations. The solution of the p-eikonal equation can be computed in similar time to shortest path distances, using a slight variation on the fast marching method (Sethian, 1996).

While the p-eikonal equations do not describe shortest path graph distances, we prove rigorously that the continuum limit of the p-eikonal equation, as the number of data points tends to infinity while p is fixed, is exactly a density weighted geodesic distance function on the underlying space (either a Euclidean domain or data manifold). Hence, the p-eikonal equation offers a robust estimation of geodesic distances in the continuum for any finite value of p. Our techniques for proving discrete to continuum convergence are quite different from existing spectral convergence results for graph Laplacians (see, e.g., (Calder et al., 2022a; Calder and García Trillos, 2022; García Trillos et al., 2020)). We use the viscosity solution machinery and the maximum principle, as in (Yuan et al., 2021; Flores et al., 2022). Our theory is also quite different from previous work on continuum limits for shortest path distances (see, e.g., (Alamgir and Von Luxburg, 2012; Bungert et al., 2022; Hwang et al., 2016; Calder et al., 2022b)) which crucially use the shortest path interpretation on the graph.

To illustrate the robustness of the p-eikonal equation, we consider applications of density weighted graph distances to data depth and semi-supervised learning. For data depth we use an approach similar to geometric medians on Riemannian manifolds (Fletcher et al., 2009). For semi-supervised learning we use a nearest neighbor classifier via the p-eikonal distance. In both applications we consider *density-weighted* distances, for which path lengths are shorter in high density regions of the graph and longer in sparse regions. This improves accuracy in semi-supervised learning and encourages the median to be placed in a high

density region of the graph in data depth problems, making the methods more robust to outliers. We test our methods on both toy and real datasets, including semi-supervised learning on MNIST, FashionMNIST and CIFAR-10. The classification results for the p-eikonal equation are uniformly better than shortest path graph distances, which we attribute to the robustness properties of the p-eikonal equation to spurious corrupted edges in real world graphs.

Using our continuum limit results, we go on to prove that p-eikonal based data depth and semi-supervised learning are asymptotically consistent. In particular, for semi-supervised learning, we take a clusterability assumption for the data and show that p-eikonal semi-supervised learning with arbitrarily few labels can recover the true labels for each cluster. The proofs of asymptotic consistency are particularly simple for graph distances, compared to the analogous results for graph Laplacian based techniques (see, e.g., (Hoffmann et al., 2022)). We also examine the role of class priors in semi-supervised learning, and show how utilizing information about the relative sizes of each class improves the asymptotic consistency results by allowing a weaker clusterability assumption. We enforce class priors by using a weighted minimum in the label decision, as was done in the volume label projection in (Calder et al., 2020a).

There is a considerable amount of related work in both data depth and semi-supervsied learning. The problem of data depth, and in general, the ordering of multivariate data, is a common problem in statistics (Barnett, 1976; Liu et al., 1999). The Tukey halfspace depth (Tukey, 1975) is one of the oldest and most well-studied notions of depths, and it has been extended to graphs (Small, 1997) and metric spaces (Carrizosa, 1996). The Tukey depth has been connected, at the continuum population level, to the solution of a non-standard Hamilton-Jacobi equation (Molina-Fructuoso and Murray, 2021). Other interesting notions of data depth include the Monge-Kantorovich depth (Chernozhukov et al., 2017), and notions of depth for curves (de Micheaux et al., 2020). Another way to define data depth is by repeatedly peeling away extremal points. Several related algorithms, including convex hull peeling, nondominated sorting, and Pareto envelope peeling, have been recently connected to viscosity solutions of partial differential equations (PDEs) in the continuum limit (Calder and Smart, 2020; Calder et al., 2014, 2015; Calder, 2016, 2017; Bou-Rabee and Morfe, 2021; Cook and Calder, 2022).

We were recently made aware of another paper (Molina-Fructuoso and Murray, 2022) that was developed in parallel with ours, and proposes to use the eikonal equation for data depth. The method in (Molina-Fructuoso and Murray, 2022) requires identifying boundary points first, and then the depth is defined as the length of a shortest density-weighted path back to the boundary. This approach, without density weighting, was also used in (Calder et al., 2022b), in combination with a method for detecting boundary points. Our approach to data depth based on the geometric median framework is much different than these works, and in particular, it does not require the *a priori* identification of boundary points to compute depth.

The problem of semi-supervised learning at low label rates has received a significant amount of attention recently, since it was pointed out in (Nadler et al., 2009) that Laplace learning (or label propagation) (Zhu et al., 2003) is ill-posed with very few labels. Many graph-based semi-supervised learning algorithms have been proposed recently at low label rates, including higher-order Laplacians (Zhou and Belkin, 2011), p-Laplacian methods

(El Alaoui et al., 2016; Kyng et al., 2015; Slepčev and Thorpe, 2019; Calder, 2018b, 2019; Flores et al., 2022), reweighted Laplacians (Shi et al., 2017; Calder and Slepčev, 2019), the centered-kernel method (Mai and Couillet, 2018a,b), volume constrained MBO (Jacobs et al., 2018), and Poisson learning (Calder et al., 2020a), and the low label rate issue has been studied theoretically in (Calder et al., 2020b). The only methods that are provably well-posed at arbitrarily low label rates are the p-Laplacian methods (Calder, 2018b; Slepčev and Thorpe, 2019) for $p > d$[1] and the Properly Weighted Laplacian (Calder and Slepčev, 2019), but neither has been shown to be asymptotically consistent at low label rates. In contrast, our results show that p-eikonal based semi-supervised learning gives well-posed, stable and informative classification results, and is asymptotically consistent, at arbitrarily low label rates and for any $p \geq 1$.

We also mention that the p-eikonal equation (1) is not new in our paper, and has already been introduced in a series of papers (Ta et al., 2009, 2010; Desquesnes et al., 2013; Desquesnes and Elmoataz, 2017). These previous works introduced the family of p-eikonal operators on graphs and presented an array of interesting applications to problems such as image segmentation, erosion, and noise removal. Our work focuses more on theoretical foundations, with our main focus being the robustness properties of the operators, and the approximation of geodesic distance in the continuum, both of which are not considered in previous work. In this sense, our results can be viewed as complementary to previous work, and provide a rigorous justification for the usefulness of the p-eikonal equation.

1.1 Outline

This paper is organized as follows. In Section 2 we study Hamilton-Jacobi equations on graphs, and introduce the p-eikonal equation. We establish our main robustness result, and then consider applications to data depth and semi-supervised learning. In Section 3 we introduce the continuum geodesic distances, and review the connection to state constrained eikonal equations. In Section 4 we prove our main discrete to continuum convergence result, showing that the p-eikonal equations recover geodesic density weighted distances in the continuum limit, for any value of $p \geq 1$. In Section 5 we use the continuum limit theory to study the asymptotic consistency of data depth and semi-supervised learning with the p-eikonal equation. Finally, in Section 6 we show the results of experiments with real data.

2. First order equations on graphs

In this section we first study the general theory of first order equations on graphs in Section 2.1 and review the graph distance function in Section 2.2. Then in Section 2.3 we introduce the p-eikonal equation, and discuss its robustness properties and computational complexity. In Section 2.4 we consider applications of the graph p-eikonal equation to data depth and semi-supervised learning. We give some toy examples in Section 2.4, and postpone experiments with real data to Section 6.

Let us first introduce some notation. Let $G = (X, W)$ be a weighted graph with vertices $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ and nonnegative edge weights $W = (w_{ij})_{i,j=1}^n$. The edge weights encode similarity between data points, with $w_{ij} \gg 0$ indicating x_i and x_j are similar, and

1. Here, d is the intrinsic dimension of the data.

$w_{ij} \approx 0$ indicating dissimilarity. We do *not* assume the weight matrix is symmetric, so in general we have $w_{ij} \neq w_{ji}$. This includes graphs such as k-nearest neighbor graphs. For first order equations, symmetry is not a main concern, since we do not require any operators to be self-adjoint, as in the case of graph Laplacians. Any zero edge weight $w_{ij} = 0$ indicates the absence of an edge from i to j. We also let $F(X)$ denote the vector space of functions $u : X \to \mathbb{R}$, and let $I_n = \{1, \dots, n\}$ denote the indices of the graph vertices. For a function $u \in F(X)$ and a vertex $x_i \in X$, we define the gradient $\nabla_X u(x_i) \in \mathbb{R}^n$ by

$$\nabla_X u(x_i) = (u(x_i) - u(x_1), u(x_i) - u(x_2), \dots, u(x_i) - u(x_n)). \tag{2}$$

For convenience, we will write $\nabla_X^j u(x_i) = u(x_i) - u(x_j)$, so that

$$\nabla_X u(x_i) = (\nabla_X^1 u(x_i), \nabla_X^2 u(x_i), \dots, \nabla_X^n u(x_i)).$$

Finally, throughout this section, we let K denote the unweighted maximum incoming degree of the graph, that is

$$K = \max_{1 \leq i \leq n} \sum_{j=1}^{n} \mathbb{1}_{w_{ji} > 0}. \tag{3}$$

2.1 General theory

We begin by developing a general theory for first order equations on graphs, and give general existence and uniqueness results. Letting $\Gamma \subset X$ denote a set of boundary or terminal nodes, a general graph PDE has the form

$$\begin{cases} H(\nabla_X u(x_i), u(x_i), x_i) = 0, & \text{if } x_i \in X \setminus \Gamma \\ u(x_i) = g(x_i), & \text{if } x_i \in \Gamma, \end{cases} \tag{4}$$

where $g : \Gamma \to \mathbb{R}$ are some prescribed boundary values. The Hamiltonian H is a function

$$H : \mathbb{R}^n \times \mathbb{R} \times X \to \mathbb{R}, \tag{5}$$

that also implicitly depends on the weight matrix W, which encodes the graph structure. It is also possible to pose a graph PDE on all of X with no boundary conditions, in the form

$$H(\nabla_X u(x_i), u(x_i), x_i) = 0 \quad \text{for all } x_i \in X. \tag{6}$$

We will write $H = H(q, z, x_i)$ in general, for $q \in \mathbb{R}^n, z \in \mathbb{R}, x_i \in X$. While we will focus on first order equations (in the sense that their continuum limits are first order PDEs), we note that this formulation of graph PDEs is very general, and contains as a subset the graph Laplacian by setting

$$H(q, z, x_i) = \sum_{j=1}^{n} w_{ij} q_j. \tag{7}$$

In this section, we establish existence and uniqueness of solutions to the graph PDE (4). Some of this analysis is similar to previous work studying PDEs on graphs, see for instance

(Manfredi et al., 2015; Calder, 2018b, 2019). Our arguments are slightly different, and cover more general cases.

Existence and uniqueness of solutions to (4) is based on a comparison principle, which allows us to compare the values of a subsolution u to a supersolution v, based on comparing their values on the boundary Γ. A subsolution $u \in F(X)$ of (4) satisfies

$$H(\nabla_X u(x_i), u(x_i), x_i) \leq 0 \quad \text{for all } x_i \in X \setminus \Gamma, \tag{8}$$

while a supersolution $v \in F(X)$ of (4) satisfies

$$H(\nabla_X v(x_i), v(x_i), x_i) \geq 0 \quad \text{for all } x_i \in X \setminus \Gamma. \tag{9}$$

Throughout this section, $\Gamma \subset X$ is fixed, and may be empty.

Definition 1. *We say that H* admits comparison *if for all $u \in F(X)$ satisfying (8) and $v \in F(X)$ satisfying (9), if $u \leq v$ on Γ then $u \leq v$ on X.*

In this section, we establish conditions under which H admits comparison. An important class of PDEs are those which are *monotone*. For vectors $p, q \in \mathbb{R}^n$, we write $p \leq q$ if $p_i \leq q_i$ for all i.

Definition 2. *We say H is* monotone *if*

$$p \leq q \text{ and } s \leq t \implies H(p, s, x) \leq H(q, t, x) \tag{10}$$

for all $x \in X$.

This definition of monotonicity is related to upwind discretizations of Hamilton-Jacobi equations, and monotone discetizations of second order equations (Sethian, 1996; Oberman, 2006). As an example, the graph Laplacian (7) is clearly monotone, since $w_{ij} \geq 0$.

Monotonicity allows us to apply maximum principle arguments to prove a comparison principle, which is based on the following observation.

Proposition 3. *Assume H is monotone and let $u, v \in F(X)$. If $u - v$ attains its maximum over X at $x_i \in X$ and $u(x_i) \geq v(x_i)$, then*

$$H(\nabla_X u(x_i), u(x_i), x_i) \geq H(\nabla_X v(x_i), v(x_i), x_i).$$

Proof We simply note that $u(x_j) - v(x_j) \leq u(x_i) - v(x_i)$ for all j, which implies that

$$u(x_i) - u(x_j) \geq v(x_i) - v(x_j) \quad \text{for all } j,$$

and so $\nabla_X u(x_i) \geq \nabla_X v(x_i)$. The result now follows from monotonicity of H ■

We can immediately prove a comparison principle when H is monotone, and one of the sub or supersolutions is strict.

Theorem 4. *Assume H is monotone. Let $u, v \in F(X)$ such that*

$$H(\nabla_X u(x_i), u(x_i), x_i) < H(\nabla_X v(x_i), v(x_i), x_i) \quad \text{for all } x_i \in X \setminus \Gamma, \tag{11}$$

and $u \leq v$ on Γ. Then $u \leq v$ on X.

Proof Let $x_i \in X$ be a point at which $u - v$ attains its maximum over X. If $x_i \in X \setminus \Gamma$, then by Proposition 3 and the assumption (11), we find that $u(x_i) < v(x_i)$. If $x_i \in \Gamma$, then $u(x_i) \leq v(x_i)$ by assumption, which completes the proof. ■

The comparison principle in Theorem 4 requires that u be a strict subsolution relative to v. The strategy to prove a true comparison principle (i.e., without the strictness, as in Definition 1) will be to make small perturbations of subsolutions (or supersolutions) to obtain the strictness required in Theorem 4. This requires that we place further assumptions on H.

Definition 5. *We say H is* proper *if there exists a strictly increasing function $\gamma : [0, \infty) \to [0, \infty)$ with $\gamma(0) = 0$ such that when $t \geq s$ we have*

$$H(q, t, x) \geq H(q, s, x) + \gamma(t - s) \tag{12}$$

for all $x \in X$ and $q \in \mathbb{R}^n$.

An example of an equation that is proper is one with a positive zeroth order term, of the form

$$H(q, z, x) = \lambda z + G(q, x),$$

where $\lambda > 0$ and $G : \mathbb{R}^n \times X \to \mathbb{R}$. In this case, $\gamma(t) = \lambda t$.

We now establish several situations where comparison holds.

Lemma 6. *Assume H is monotone. Then H admits comparison if any of the following hold.*

(i) *H is proper.*

(ii) *$H = H(q, x)$, $q \mapsto H(q, x)$ is convex, and there exists $\varphi \in F(X)$ and $\lambda > 0$ such that*

$$H(\nabla_X \varphi(x_i), x_i) + \lambda \leq 0 \quad \text{for all} \quad x_i \in X \setminus \Gamma. \tag{13}$$

(iii) *$H(q, z, x) = G(q) - f(x)$, where $f > 0$ on X, and G is positively p-homogeneous for $p > 0$.*

Proof Let u satisfy (8) and v satisfy (9), and assume that $u \leq v$ on Γ. In each case we will produce a perturbation u_ε of u satisfying $u_\varepsilon \leq v$ on Γ, $H(\nabla_X u_\varepsilon, u, x) < 0$, and $u_\varepsilon \to u$ as $\varepsilon \to 0$. Then by Theorem 4 we have $u_\varepsilon \leq v$ and sending $\varepsilon \to 0$ completes the proof.

(i) We set $u_\varepsilon = u - \varepsilon$ and use the fact that H is proper to get the strict subsolution condition.

(ii) We set $u_\varepsilon = (1 - \varepsilon)u + \varepsilon\varphi$. We can shift φ by a constant, if necessary, so that $\varphi - u \leq 0$, and so $u_\varepsilon \leq u$. Since $q \mapsto H(q, x)$ is convex, we have

$$\begin{aligned} H(\nabla_X u_\varepsilon(x_i), x_i) &= H((1 - \varepsilon)\nabla_X u(x_i) + \varepsilon \nabla_X \varphi(x_i), x_i) \\ &\leq (1 - \varepsilon) H(\nabla_X u(x_i), x_i) + \varepsilon H(\nabla_X \varphi(x_i), x_i) \leq -\lambda\varepsilon, \end{aligned}$$

for all $x_i \in X \setminus \Gamma$.

(iii) Define $u_\varepsilon = (1-\varepsilon)u + \varepsilon \min_X u$. Then $u_\varepsilon \leq u$. Since G is positively p-homogeneous we have $G(aq) = |a|^p G(q)$ for all $a \in \mathbb{R}$ and $q \in \mathbb{R}^n$, and so

$$G(\nabla_X u_\varepsilon(x_i)) = G((1-\varepsilon)\nabla_X u(x_i)) = (1-\varepsilon)^p G(\nabla_X u(x_i)) \leq (1-\varepsilon)^p f(x_i).$$

Hence, we have

$$G(\nabla_X u_\varepsilon(x_i)) - f(x_i) \leq -\left(1 - (1-\varepsilon)^p\right) f(x_i) < 0.$$

■

If H admits comparison, then we can prove existence of a solution to (4) using the Perron method. We summarize this in the following result.

Theorem 7. *Assume H is monotone, continuous in p and z, and admits comparison. Assume there exists $\varphi, \psi \in F(X)$ such that $\psi \geq \varphi = g$ on Γ and for $x_i \in X \setminus \Gamma$*

$$H(\nabla_X \varphi(x_i), \varphi(x_i), x_i) \leq 0 \quad \textit{and} \quad H(\nabla_X \psi(x_i), \psi(x_i), x_i) \geq 0.$$

Then there exists a unique solution $u \in F(X)$ of (4) and $\varphi \leq u \leq \psi$.

The proof of Theorem 7 is very similar to existing results (e.g., Theorem 4 of (Calder, 2018b)). We include the proof in Appendix B for reference.

Remark 8. *Notice that none of the results in this section have required graph connectivity, which is a common assumption in the analysis of PDEs on graphs. Normally, graph connectivity is used in a* path to the boundary *argument to establish a comparison principle (see, e.g., (Manfredi et al., 2015; Calder, 2018b, 2019)). Our arguments do not require graph connectivity to establish comparison. The one place connectivity requirements may appear is in the construction of the super and subsolutions φ and ψ in the Perron method in Theorem 7.*

2.2 Graph distance functions

The graph distance $d_G : X \times X \to \mathbb{R}$ is defined by

$$d_G(x_i, x_j) = \min_{m \geq 1} \min_{\tau \in I_n^m} \left\{ w_{i,\tau_1}^{-1} + \sum_{i=1}^{m-1} w_{\tau_i,\tau_{i+1}}^{-1} + w_{\tau_m,j}^{-1} \right\}, \tag{14}$$

where we recall that $I_n = \{1, \ldots, n\}$, and $I_n^m = (I_n)^m$. We use the interpretation that $w_{ij}^{-1} = \infty$ whenever $w_{ij} = 0$, which implicitly restricts the feasible paths to follow edges in the graph and to connect x_i to x_j.

Definition 9. *We say that the graph G is connected if $d_G(x_i, x_j) < \infty$ for all $x_i, x_j \in X$.*

We also define the graph distance to a set $\Gamma \subset X$ as follows

$$d_G(x_i, \Gamma) = \min_{x_j \in \Gamma} d_G(x_i, x_j).$$

We recall that the graph distance function satisfies a certain graph eikonal equation. The result is well-known (see, e.g., Lemma 3 of (Bungert et al., 2022)), but usually stated for symmetric graphs, so we will sketch a proof for completeness.

Lemma 10 ((Bungert et al., 2022)). *Assume G is connected and let $\Gamma \subset X$. Then the graph distance function $u(x) := d_G(x, \Gamma)$ is the unique solution of the graph eikonal equation*

$$\max_{x_j \in X} w_{ji}(u(x_i) - u(x_j)) = 1 \quad \text{for all } x_i \in X \setminus \Gamma, \tag{15}$$

satisfying $u(x_i) = 0$ for $x_i \in \Gamma$.

Remark 11. *We call* (15) *the* graph eikonal equation*, since its solution is a distance function, in the same way that the continuum eikonal equation (see Section 3) represents continuum path distances. In the notation of Section 2.1, the graph eikonal equation corresponds to the* monotone *Hamiltonian $H(q, x_i) = \max_{1 \le j \le n} w_{ji} q_j - 1$.*

In terms of computational complexity, the solution of (15) *can be computed with Dijkstra's algorithm in $\mathcal{O}(nK \log(n))$ time, where we recall K is the maximum (unweighted) degree of any node in the graph, defined in* (3)*.*

Proof [Proof of Lemma 10] The main idea of the proof is to use the fact that u satisfies the dynamic programming principle

$$u(x_i) = \min_{x_j \in X} (u(x_j) + w_{ji}^{-1}). \tag{16}$$

Since the graph is connected, there exists some j with $w_{ji} > 0$, and both $u(x_j)$ and $u(x_i)$ are finite. We can rearrange this to obtain

$$\max_{x_j \in X} (u(x_i) - u(x_j) - w_{ji}^{-1}) = 0.$$

Since the max is zero, we can multiply by w_{ji} inside the brackets above and rearrange to obtain the result. To prove uniqueness, we can run the proof in the opposite direction, showing that any solution of (15) satisfies the dynamic programming principle (16), and is thus the graph distance function $d_G(\cdot, \Gamma)$. ■

It is common to consider density weighted distances in data science and machine learning applications. This allows us to make it more expensive for paths to travel through sparse regions in space, and less expensive to travel within dense regions. This makes points within clusters closer together, while driving points in different clusters further apart, which is useful for cluster and semi-supervised learning.

In the context of the graph eikonal equation (15), density weighting can be introduced by solving the equation with a right hand side, of the form

$$\max_{x_j \in X} w_{ji}(u(x_i) - u(x_j)) = f(x_i) \quad \text{for all } x_i \in X \setminus \Gamma. \tag{17}$$

One can choose, for example, $f(x_i) = \hat{\rho}(x_i)^{-\alpha}$ for $\alpha \ge 0$, where $\hat{\rho} : X \to \mathbb{R}$ is any density estimator (say, a kernel density estimator or a k-nearest neighbor estimator), and α is a tunable parameter. Since we did not assume the graph was connected in Lemma 10, we can apply the lemma to (17) with the graph weights $\overline{w}_{ij} = f(x_j)^{-1} w_{ij}$ to obtain that the

solution of (17) subject to $u(x_i) = 0$ for $x_i \in \Gamma$ corresponds to the density weighted graph distance

$$d_{G,f}(x_i, x_j) := \min_{m \geq 1} \min_{\tau \in I_n^m} \left\{ w_{i,\tau_1}^{-1} f(x_{\tau_1}) + \sum_{i=1}^{m-1} w_{\tau_i,\tau_{i+1}}^{-1} f(x_{\tau_{i+1}}) + w_{\tau_m,j}^{-1} f(x_{\tau_j}) \right\}. \tag{18}$$

When $f = \hat{\rho}^{-\alpha}$ with $\alpha \geq 0$, the reweighted equation (17) makes it more expensive for paths to travel through regions where the density, $\hat{\rho}$, is low, and less expensive where the density is high. Of course, choosing $\alpha \leq 0$ has the opposite effect.

2.2.1 Sensitivity to noise

We mention that the graph eikonal equation (17) is highly sensitive to corruption in the weight matrix W used to construct the graph. Indeed, we can see this quite easily from the distance function interpretation, since adding a single spurious edge between two distant nodes in a graph creates a *short-cut* that drastically changes the distance function. Thus, while the graph eikonal equation (17) does indeed approximate geodesic distances on the underlying data manifold well (see, e.g., (Hwang et al., 2016)), the graph distance lacks robustness to noise and other corruptions. To illustrate this, we refer to Figure 1a, which shows how drastically the graph distance function can change with the addition of a few spurious edges in the graph. The graph is a simple unweighted proximity graph on $n = 20000$ uniformly distributed random variables on the unit ball. Points within distance $\varepsilon = 0.05$ are connected by an edge with edge weight of 1, and the boundary set Γ is chosen to be all points within distance ε of the boundary of the ball. From left to right in Figure 1a, we add 0, 10, 20, and 50 corrupted edges at random, and show the resulting distance functions to the boundary.

2.3 The p-eikonal equation

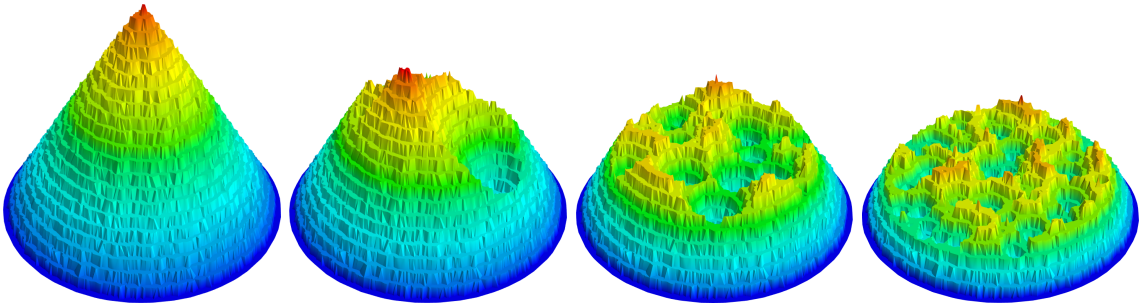
The issue with lack of robustness of the graph eikonal equation (17) stems from the form of the *max* in the operator, which means its value is highly sensitive to a single outlying edge weight. We introduce here the *p-eikonal equation* on a graph, which uses information from all neighbors, and as we will show below, gives a more robust distance function on a graph. For $p > 0$, we define the p-eikonal operator $\mathcal{A}_{G,p} : F(X) \to F(X)$ by

$$\mathcal{A}_{G,p} u(x_i) = \sum_{j=1}^{n} w_{ji} (u(x_i) - u(x_j))_+^p, \tag{19}$$

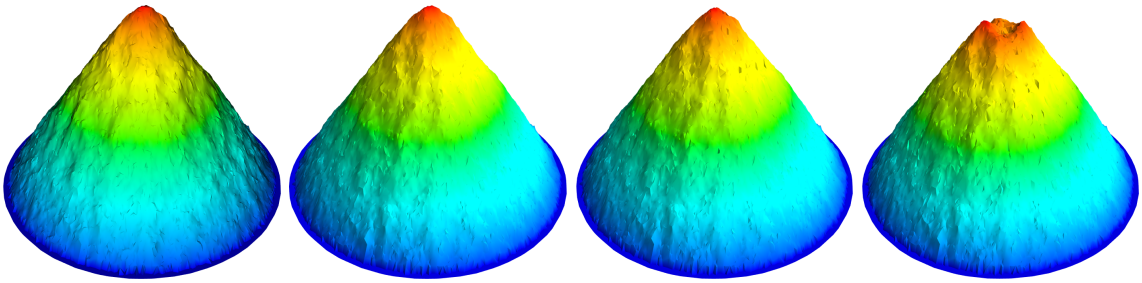
where $a_+ := \max\{a, 0\}$ is the positive part. For $\Gamma \subset X$ and $f \in F(X)$, we consider the p-eikonal equation

$$\begin{cases} \mathcal{A}_{G,p} u = f, & \text{in } X \setminus \Gamma \\ u = 0, & \text{on } \Gamma. \end{cases} \tag{20}$$

We show in Figure 1b the robustness experiment described in the last section with the p-eikonal equation with $p = 1$. The p-eikonal equation is clearly more robust to the additional corrupted edges in the graph. After some preliminary results, we prove in Theorem 14,



(a) Graph distance function with corrupted edges



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

Figure 1: Robustness of graph-distance functions compared to the p-eikonal equation under random corruptions of edges in the graph. We computed each distance function on an unweighted proximity graph over $n = 20000$ uniformly distributed random variables on the unit ball with graph connectivity length scale $\varepsilon = 0.05$. The boundary points Γ were chosen to be all points within ε of the boundary of the unit ball, so the distance function gives a notion of data depth. From left to right we added an increasing number of corrupted edges (0, 10, 50, and 200) with edge weight $w_{ij} = 1$. We see the solution of the p-eikonal equation is far more robust under the addition of corrupted edges.

below, a robustness estimate for the p-eikonal equation that explains the experimental results in Figure 1b.

We first use the theory from Section 2.1 to establish that (20) is well-posed. For $q \geq 0$ we denote by G^q the graph $G^q = (X, W^q)$ with weights $W^q = (w_{ij}^q)_{i,j=1}^n$. We interpret $0^0 = 0$ so that G^0 is the unweighted graph with the same edges as G.

Theorem 12 (Well-posedness)**.** *Let $p > 0$ and $f > 0$. If G is connected, then* (20) *has a unique solution $u \in F(X)$, and*

$$K^{-\frac{1}{p}} \left(\min_X f^{\frac{1}{p}} \right) d_{G^{\frac{1}{p}}}(x_i, \Gamma) \leq u(x_i) \leq \left(\max_X f^{\frac{1}{p}} \right) d_{G^{\frac{1}{p}}}(x_i, \Gamma). \tag{21}$$

Proof In the notation of Section 2.1, the p-eikonal equation (20) corresponds to the Hamiltonian

$$H(q, x_i) = \sum_{j=1}^n w_{ji} (q_j)_+^p - f(x_i).$$

This Hamiltonian is monotone, and positively p-homogeneous (and also convex in q when $p \geq 1$). Thus, (20) admits comparison by Lemma 6. Hence, existence follows from the Perron method (Theorem 7), provided we can exhibit a subsolution φ and supersolution ψ with $\varphi = 0 \leq \psi$ on Γ. We can take $\varphi = 0$, but we will construct a larger subsolution to prove the bound (21).

For $c > 0$ to be determined, let us define

$$\varphi(x_i) = c d_{G^{\frac{1}{p}}}(x_i, \Gamma).$$

By Lemma 10, φ solves the graph eikonal equation

$$\max_{x_j \in X} w_{ji}^{\frac{1}{p}} (\varphi(x_i) - \varphi(x_j))_+ = c.$$

Since the right hand side c is positive, we can trivially add the positive part above. Then we have

$$\begin{aligned}
\mathcal{A}_{G,p}\varphi(x_i) &= \sum_{j=1}^{n} w_{ji}(\varphi(x_i) - \varphi(x_j))_+^p \\
&\leq K \max_{x_j \in X} w_{ji}(\varphi(x_i) - \varphi(x_j))_+^p \\
&= K \left(\max_{x_j \in X} w_{ji}^{\frac{1}{p}} (\varphi(x_i) - \varphi(x_j))_+ \right)^p \\
&= K c^p.
\end{aligned}$$

Setting $c = K^{-\frac{1}{p}} \min_X f^{\frac{1}{p}}$, we have $\mathcal{A}_{G,p}\varphi \leq f$ on $X \setminus \Gamma$, which proves the subsolution condition. We likewise define

$$\psi(x_i) = C d_{G^{\frac{1}{p}}}(x_i, \Gamma),$$

and use a similar argument to find that a choice of $C = \max_X f^{\frac{1}{p}}$ yields the supersolution condition. ∎

Remark 13. *Let u_p for $p > 0$ denote the solution of* (20)*, which exists and is unique due to Theorem 12. By* (21) *we see that $u_p \to d_{G^0}(\cdot, \Gamma)$ as $p \to \infty$. Thus, the $p \to \infty$ limit of the p-eikonal equation recovers the unweighted graph distance. By a similar argument, the solution of $\mathcal{A}_{G^p,p} u_p = f^p$ will satisfy $u_p \to u$ as $p \to \infty$, where u is the solution of the graph eikonal equation* (17)*.*

2.3.1 Robustness to noise

We now turn to the question of robustness of the p-eikonal equation to graph perturbations. We consider a perturbation $\widetilde{W} = W + \delta W$, where the only requirement is that the pertrubed matrix $\widetilde{W}$ is a valid similarity matrix (i.e., has nonnegative entries), and that the graph remains connected after the pertrubation, so that the p-eikonal equation has a unique solution. This allow the addition or deletion of edges, or simply the modification of existing edges. We define $A_+ = \max\{A, 0\}$ and $A_- = \min\{A, 0\}$ for a matrix A, where the minimum and maximum are pointwise.

Theorem 14. *Assume $G = (X, W)$ is connected. Let $\delta W \in \mathbb{R}^{n\times n}$ such that $\widetilde{W} := W + \delta W \geq 0$ and $\widetilde{G} := (X, \widetilde{W})$ is connected. Let $\Gamma \subset X$, $f \in F(X)$ with $f > 0$ and let $u, \widetilde{u} \in F(X)$ satisfy*

$$\begin{cases} \mathcal{A}_{\widetilde{G},p}\widetilde{u}(x_i) = \mathcal{A}_{G,p}u(x_i) = f(x_i), & \text{if } x_i \in X \setminus \Gamma \\ \widetilde{u}(x_i) = u(x_i) = 0, & \text{if } x_i \in \Gamma. \end{cases} \tag{22}$$

Then for all $x_i \in X$ we have

$$-\left(\max_{X\setminus\Gamma} \frac{\mathcal{A}_{\delta G_-,p}\widetilde{u}}{f}\right)^{\frac{1}{p}} \leq \frac{u(x_i) - \widetilde{u}(x_i)}{\min\{u(x_i), \widetilde{u}(x_i)\}} \leq \left(\max_{X\setminus\Gamma} \frac{\mathcal{A}_{\delta G_+,p}u}{f}\right)^{\frac{1}{p}}, \tag{23}$$

where $\delta G_\pm = (X, \pm\delta W_\pm)$.

Proof For notational simplicity, let us set

$$\delta = \max_{X\setminus\Gamma} \frac{\mathcal{A}_{\delta G_+,p}u}{f}.$$

Let $G_+ = (X, W + \delta W_+)$ and let u_- denote the solution of $A_{G_+,p}u_+ = f$ subject to $u_+ = 0$ on Γ. Notice we have

$$\mathcal{A}_{G,p}u = f = \mathcal{A}_{G_+,p}u_+ \geq \mathcal{A}_{G,p}u_+,$$

and so it follows from comparison that $u_+ \leq u$. A similar argument shows that $u_+ \leq \widetilde{u}$. We now compute

$$\frac{\mathcal{A}_{G_+,p}u(x_i)}{\mathcal{A}_{G_+,p}u_+(x_i)} \leq \frac{f(x_i) + \mathcal{A}_{\delta G_+,p}u(x_i)}{f(x_i)} \leq 1 + \max_{X\setminus\Gamma} \frac{\mathcal{A}_{\delta G_+,p}u}{f} = 1 + \delta$$

for all $x_i \in X \setminus \Gamma$. Therefore $\mathcal{A}_{G_+,p}u \leq \mathcal{A}_{G_+,p}((1+\delta)^{\frac{1}{p}}u_+)$ on $X \setminus \Gamma$. By the comparison principle we have

$$u \leq (1+\delta)^{\frac{1}{p}}u_+ \leq \widetilde{u} + \delta^{\frac{1}{p}} \min\{u, \widetilde{u}\},$$

as $u_+ \leq \min\{u, \widetilde{u}\}$. Therefore $u - \widetilde{u} \leq \delta^{\frac{1}{p}} \min\{u, \widetilde{u}\}$, which completes the proof of the upper bound in (23).

To prove the lower bound, we simply swap u and $\widetilde{u}$ in the argument above, and use that $W = \widetilde{W} - \delta W$. ∎

Remark 15. *Theorem 14 controls the relative error between u and $\widetilde{u}$. We note, in particular, that the dependence on p shows that $p = 1$ offers the greatest robustness, and as $p \to \infty$ we lose the robustness completely. We note that there are several ways we can reformulate Theorem 14. First, let us define the* upwind 1*-norm of a matrix, relative to the function $u \in F(X)$, by*

$$\|A\|_{u,1} = \max_{1\leq j\leq n} \sum_{i=1}^{n} |A_{ij}| \mathbb{1}_{u(x_j) > u(x_i)}.$$

We note that $\|A\|_{u,1} \leq \|A\|_1$*, where* $\|A\|_1 = \max_{1\leq j\leq n}\sum_{i=1}^n |A_{ij}|$ *is the usual* 1*-norm. The norm* $\|\delta W\|_{u,1}$ *measures the maximum amount of corruption among the incoming edges of any node from directions where* u *is smaller (the* upwind *direction). Then we compute*

$$\mathcal{A}_{\delta G_\pm,p}u(x_i) = \pm\sum_{j=1}^n (\delta w_{ji})_\pm (u(x_i)-u(x_j))_+^p \leq u(x_i)^p \sum_{j=1}^n |\delta w_{ji}| \mathbb{1}_{u(x_i)>u(x_j)} \leq u(x_i)^p \|\delta W\|_{u,1}.$$

Thus, for example, the upper bound in Theorem 14 implies that

$$\frac{u(x_i)-\widetilde{u}(x_i)}{\min\{u(x_i),\widetilde{u}(x_i)\}} \leq \left(\max_{X\setminus\Gamma} \frac{u}{f^{\frac{1}{p}}}\right) \|\delta W\|_{u,1}.$$

Finally, using the upper bound in Theorem 12 we obtain

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

where $0 < f_{min} \leq f \leq f_{max}$ *and* $C = \max_{x_i\in X} d_{G^{\frac{1}{p}}}(x_i,\Gamma)$*. A similar statement holds for the lower bound in thm:robust.*

2.3.2 Computational complexity

The p-eikonal equation (20) can be solved in a similar computational time as the graph eikonal equation (17) using the fast marching method (Sethian, 1996) on a graph. The solution of (20) via fast marching requires repeatedly solving the equation

$$\sum_{j=1}^n w_{ji}(t-s_j)_+^p = a, \tag{24}$$

for the unknown t, given s_j, $j = 1,\ldots,n$, and a. Of course, only the s_j with $w_{ji} > 0$ need to be considered. Since the left hand side is increasing in t, the equation can be solved with a bisection search for any $p > 0$. Using a tolerance of δ, the complexity of solving the scheme (24) with a bisection search is $\mathcal{O}(K\log(\delta^{-1}))$, where K is the maximum unweighted degree of the graph defined in (3).

When $p = 1$, we can in fact solve the scheme (24) explicitly without a bisection search. We first sort the s_j in ascending order (and relabel the w_{ji} in the same order), and then note that the solution t will have the form

$$t = t_m := \frac{a+\sum_{j=1}^m w_{ji}s_j}{\sum_{j=1}^m w_{ji}},$$

for some $m \leq n$. We can compute all the t_m recursively in $\mathcal{O}(K\log(K))$ time, and simply check which is correct, yielding $\mathcal{O}(K\log(K))$ complexity for solving (24) when $p = 1$. A similar observation can be made for $p = 2$, except that $t = t_m$ will be the solution of a quadratic equation.

The fast marching method visits each node in the graph exactly once, in order of increasing values of the solution $u(x_i)$. When each node is visited, the scheme (24) is solved

at all neighbors of the node. Each time the scheme is solved, a heap[2] of size at most n is updated, which takes $\log(n)$ time. Thus the fast marching method takes $\mathcal{O}(nK^2 \log(K) \log(n))$ computational time for $p = 1$ or $p = 2$, and $\mathcal{O}(nK^2 \log(\delta^{-1}) \log(n))$ time for other positive values of p, where δ is the bisection solver tolerance. In our implementation, of the method, we use the exact solution of the scheme for $p = 1$, and the bisection search for all $p > 1$ (i.e., we did not implement the quadratic method described above for $p = 2$, since we found it did not improve over the bisection search).

2.3.3 Shortest paths

While the solution of the p-eikonal equation (20) does not represent a true distance function on the graph, as the eikonal equation (17) does, we can still construct a notion of a *shortest path* from any $x_i \in X \setminus \Gamma$ back to the set Γ, by descending on u as quickly as possible. In particular, given the solution u of (20) and an initial point $x_{i_0} \in X \setminus \Gamma$, we select the next point, for $k \geq 0$, to satisfy

$$x_{i_{k+1}} \in \operatorname*{argmin}_{\substack{x_j \in X \\ w_{j,i_k} > 0}} u(x_j).$$

In other words, the next point is the neighbor of x_{i_k} with the smallest value of u, which is the "closest" to Γ. Provided $f > 0$ and $x_{i_k} \notin \Gamma$, there must exist a neighbor with a strictly smaller value for u, otherwise we would have $\mathcal{A}_{G,p} u(x_{i_k}) = 0 < f(x_{i_k})$, which contradicts that u solves the p-eikonal equation (20). Thus, the path chosen in this way is strictly decreasing in u, that is

$$u(x_{i_0}) > u(x_{i_1}) > u(x_{i_2}) > \cdots .$$

This guarantees that the path can never visit a node twice, and will eventually terminate at a point $x_{i_T} \in \Gamma$ after some number of steps, T. The shortest paths computed in this way are shown in red in the data depth experiments in Figures 2 and 3. We also use this method to compute the shortest paths through real data in Section 6.

2.4 Applications

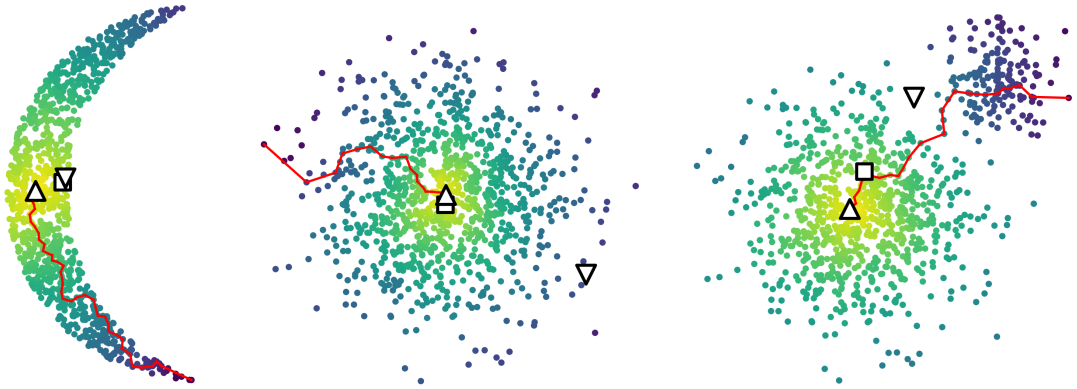
The solution of the p-eikonal equation (20), while not a true graph distance function, gives us a type of approximate distance that is useful for data depth and semi-supervised learning. We discuss these applications initially in this section, and show the results of some experiments on toy datasets. We postpone experiments with real data to Section 6.

Given a set $\Gamma \subset X$ and a density estimation $\hat{\rho} : X \to \mathbb{R}$, we consider solving the density reweighted p-eikonal equation

$$\begin{cases} \mathcal{A}_{G,p} u = \hat{\rho}^{-\alpha}, & \text{in } X \setminus \Gamma \\ u = 0, & \text{on } \Gamma, \end{cases} \tag{25}$$

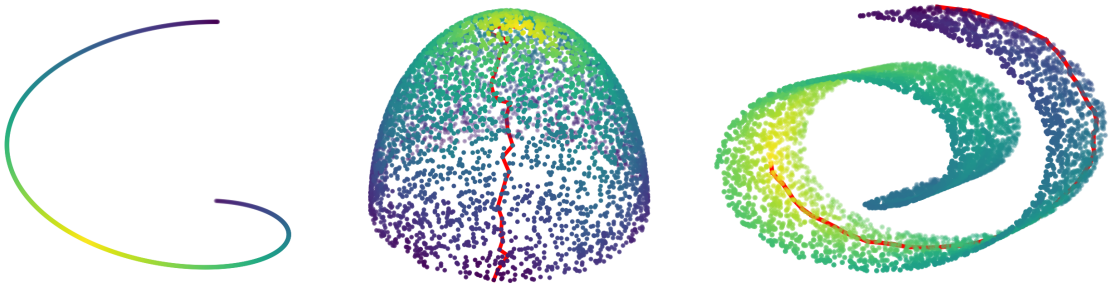
where the exponent α is a tunable parameter. We denote the solution of (25) by $D_\Gamma^{p,\alpha}(x) = u(x)$. When $\Gamma = \{x\}$ is a single point we write $D_x^{p,\alpha}$.

2. The heap stores the current best guesses of $u(x_i)$ for nodes x_i that have not been finalized/visited yet. At each iteration of fast marching, we need to retrieve the node with smallest best guess, which can be done in $\mathcal{O}(\log(n))$ time with a heap data structure. When updating the scheme at all neighbors, the heap needs to be updated, also taking $\mathcal{O}(\log(n))$ time.



(a) Moon (b) Gaussian (c) Gaussian mixture

Figure 2: The p-eikonal medians and depth on 2D toy datasets with $p = 1$ (see Section 2.4.1 for definitions). The medians are shown for different values of the density parameter α in (25) with $\alpha = -1$ ($\triangledown$), $\alpha = 0$ ($\square$) and the $\alpha = 1$ ($\triangle$), while the points are colored by the $\alpha = 1$ data depth. We also show the shortest path from the shallowest point to the deepest point in red. We only recommend $\alpha \geq 0$ in all our applications; we have shown $\alpha = -1$ just to illustrate how reverse density weighting affects the median computation (in this case, it prefers placing the median in sparse regions of the graph).



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

Figure 3: The p-eikonal data depth on 3D toy datasets sampled from manifolds embedded in $\mathbb{R}^3$. We use $p = 1$ and $\alpha = 1$. We note that the swiss roll is more dense on one end than the other, which explains why the depth is not symmetric along the length of the roll.

2.4.1 Data depth

We can approach data depth through the framework of the geometric median. Let us recall that for a collection of points $x_1, \ldots, x_n$ in $\mathbb{R}^d$, the *geometric median* x_* is defined by

$$x_* \in \operatorname*{argmin}_{x \in \mathbb{R}^d} \sum_{i=1}^{n} |x_i - x|.$$

The geometric median generalizes the 1-dimensional median, and inherits many of its robustness properties (its breakdown point is also 0.5, for example). Given the notion of depth $D_x^{p,\alpha}$, we define the p-eikonal median $x_{p,\alpha}$ by

$$x_{p,\alpha} \in \operatorname*{argmin}_{x_j \in X} \sum_{x_i \in X} D_{x_j}^{p,\alpha}(x_i). \tag{26}$$

In practice, we approximate the median by restricting $x_j \in \hat{X} \subset X$, where $\hat{X}$ is a much smaller subset of X chosen at random. In all our experiments we take $\hat{X}$ to have 5% of the points in X.

Once we have computed the median $x_{p,\alpha}$, we obtain a notion of data depth via the distance to the median

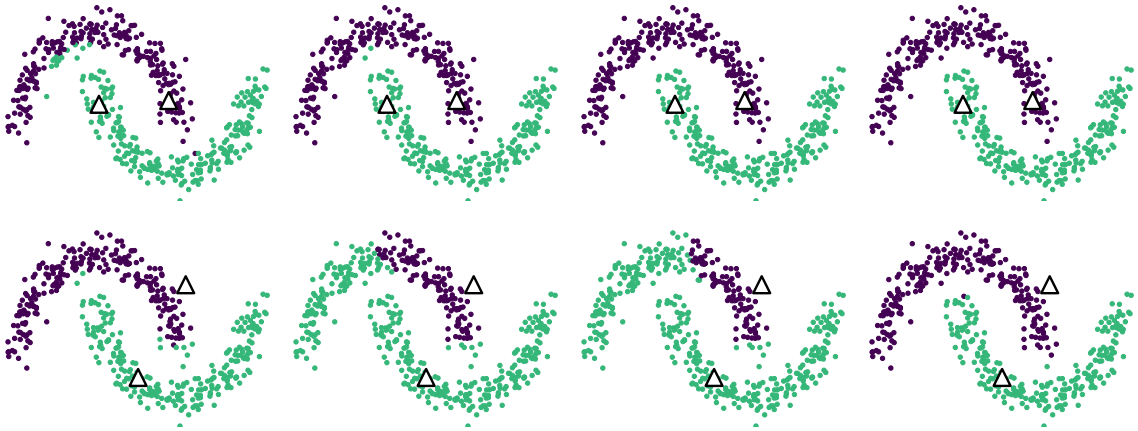
$$\operatorname{depth}_{p,\alpha}(x) = \max_{y \in X} D_{x_{p,\alpha}}^{p,\alpha}(y) - D_{x_{p,\alpha}}^{p,\alpha}(x).$$

Figure 2 gives an example of the medians and depths for different toy datasets in 2 dimensions, and for $\alpha \in \{-1, 0, 1\}$. We use $p = 1$ in all experiments, and color the point cloud by the $\alpha = 1$ depth. We can see that the $\alpha = 1$ median outperforms the other weighting choices. In particular, in the Gaussian mixture example, the $\alpha = 1$ median is completely insensitive to the addition of the outlying cluster, which has $\frac{1}{6}$ of the points in the main cluster. This insensitivy is desirable in robust statistics, as it ensures that the p-eikonal median, like its one dimensional counterpart, is robust in the presence of noise, such as outliers (compared, for example, to the mean, which lacks such robustness). We show in Figure 3 example of the p-eikonal median and depth on point clouds sampled from submanifolds of $\mathbb{R}^3$. In this case we just show the $\alpha = 1$ depth. In all images we also show the shortest path, computed as described in Section 2.3.3, from the shallowest to the deepest point.

Let us remark briefly that the density weighting with $\alpha > 0$ encourages the median to be placed in regions of high density, since path lengths are shorter here. In contrast, taking $\alpha < 0$ encourages the median to be in regions of low density. We do not recommend using $\alpha < 0$ in data depth (or in semi-supervised learning). We also remark that in Figure 3, the depth on the swiss roll is not symmetric along the length of the roll. This is to be expected with density weighting, since the swiss roll is more dense near one end of the roll (near the origin) and less dense on the other end. We postpone examples of the p-eikonal depth on real data to Section 6.

2.4.2 Semi-supervised learning

Given the approximate distances $D_\Gamma^{p,\alpha}$ we can perform semi-supervised learning with a nearest neighbor approach. Suppose we have k classes, and for each class $j = 1, \ldots, k$, we



(e) $\alpha = -1$ (f) $\alpha = 0$ (g) $\alpha = 1$ (h) $\alpha = 1$ with class priors

Figure 4: Example of semi-supervised learning with the density weighted p-eikonal equation on the two moons dataset. The $\triangle$ markers give the locations of the two labeled points in each example. We show different choices of density reweighting, and the addition of class priors on the right side.

are provided some labeled nodes $\Gamma_j \subset X$. The label prediction ℓ_i for an unlabeled node $x_i \notin \Gamma_j$ for any j, is the label of the closest labeled node, under the distance $D_{\Gamma}^{p,\alpha}$, that is

$$\ell_i = \operatorname*{argmin}_{1 \leq j \leq k} D_{\Gamma_j}^{p,\alpha}(x_i). \tag{27}$$

Semi-supervised learning with the p-eikonal equation thus requires solving k separate p-eikonal equations, which is similar to the one-vs-rest approach in machine learning for producing a multi-class classifier out of a binary one.

As we shall see in our analysis later in Section 5, distance-based classifiers can be highly sensitive to the geometry of the clusters, even with appropriate density weighting. In such cases, we can improve the accuracy of the classifier by incorporating information about *class priors*, so that the classifier predicts the correct proportion of nodes in each class. To do this, we follow (Calder et al., 2020a) and modify the label decision with the addition of positive weights $s_1, \ldots, s_k$ so that the new label decision is

$$\ell_i = \operatorname*{argmin}_{1 \leq j \leq k} \left\{ s_j D_{\Gamma_j}^{p,\alpha}(x_i) \right\}. \tag{28}$$

By increasing or decreasing the weights s_j, we can increase or decrease the number of nodes predicted in each class. The weights s_j can be adjusted incrementally until class balancing is achieved. We do this with the volume constrained label projection method from (Calder et al., 2020a).

As a preliminary toy example, we consider classification of the two-moons dataset in Figure 4. The two rows in the figure correspond to different choices of the labeled nodes. In each case we take one label per class and indicate its position with a $\triangle$. In the first row, the training nodes are both inliers in their respective clusters, and all choices of weighting

exponents α give good classification, and the addition of class priors is not needed. In the second row, the training point for the upper half of the moon is an outlier for that cluster, and the lower cluster leaks over significantly for $\alpha = 0, 1$. We see on the right that this issue can be corrected with the addition of class prior information, to enforce the predicted classes to have the same size. It is also interesting to note in the second row that the reverse density weighting $\alpha = -1$ produces the correct classification without class priors. This is because the reverse density weighting brings the outlying training point closer to its cluster. In general, when we do not expect training points to be outliers, we do not recommend reverse density weighting in semi-supervised learning (and we do not observe good results with reverse density weighting with real data). In Section 6 we present more in depth results with semi-supervised learning on real data. We remark here that other types of perturbation like missing data or undersampled graphs maybe an interesting subject for future investigations.

3. State-constrained eikonal equations

The continuum limit of the graph p-eikonal equation is a PDE called the state-constrained eikonal equation, which has the form[3]

$$\begin{cases} |\nabla u| = f, & \text{in } \Omega \setminus \Gamma \\ u = 0, & \text{on } \Gamma. \end{cases} \tag{29}$$

Here, u is a function $u : \Omega \to \mathbb{R}$ and ∇u denotes the gradient of u, which is the vector of partial derivatives $\nabla u = (u_{x_1}, u_{x_2}, \ldots, u_{x_d})$ where u_{x_i} is the partial derivative in the i^{th} coordinate. The notation $|\nabla u|$ is the Euclidean norm of ∇u. The equation is *state-constrained* because, as we shall see below, the solution represents a geodesic distance function to Γ, and the set Ω constrains the geodesic paths (e.g., the state).

Remark 16. *For the reader who is unfamiliar with PDEs, let us give an example of a solution to the state constrained equation* (29) *in the special case that* $d = 1$, $f \equiv 1$,

$$\Omega = (-2, 2) \quad \textit{and} \quad \Gamma = \{-1, 1\}.$$

In this case, the solution of (29) *should satisfy* $|u'(x)| = 1$ *for* $x \neq \pm 1$, *and* $u(\pm 1) = 0$. *However, there is no continuously differentiable function with these properties. Indeed, any such function would have a critical point in the interval* $(-1, 1)$, *which contradicts the equation* $|u'(x)| = 1$. *This well-known issue with first order Hamilton-Jacobi equations led to the development of a weak notion of solution for PDEs known as the* viscosity solution *(Crandall and Lions, 1983; Crandall et al., 1984), which we define below for the state constrained problem. In this example, the viscosity solution is given by*

$$u(x) = \min\{|x + 1|, |x - 1|\}, \tag{30}$$

3. We note that the f in (29) is not the same as the f on the right hand side of the p-eikonal graph equation (20), as we have absorbed the density and exponent p into a generic right hand side (see (48) for the precise form of the continuum limit). It is more convenient in this section to treat the general state constrained equation with any right hand side.

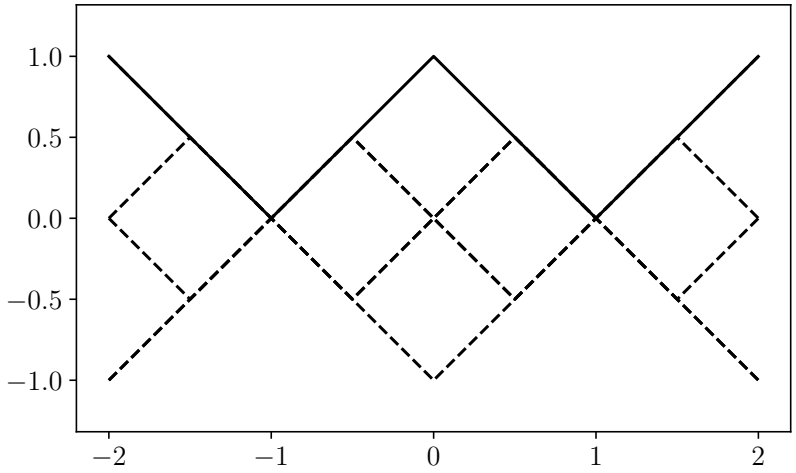


Figure 5: The viscosity solution discussed in Remark 16 is shown as a solid line, while some other Lipschitz functions that solve the state-constrained PDE at all points of differentiability are shown as dotted lines.

and is depicted in Figure 5. Notice that $u(x)$ *is exactly the distance function* $u(x) = \min_{y\in\Gamma} |x-y|$ *to the set* Γ*, which is also explained below in this section.*

We also remark that the solution u *given above satisfies* $|u'(x)| = 1$ *at all points* $x \in \Omega\setminus\Gamma$ *except the point* $x = 0$ *where the function is not differentiable. Clearly there are other such functions that satisfy the equation at all points of differentiability; for example, the function* $v(x) = -u(x)$*, or the function*

$$w(x) = \min\{|x+1|, |x-1|, |x|\}.$$

In fact, if we merely look for a Lipschitz continuous function u *satisfying* $|u'(x)| = 1$ *at all points of differentiability and* $u(\pm 1) = 0$*, then there are infinitely many such functions (see Figure 5 for a depiction of some of them). The notion of viscosity solution selects a particular solution out of this infinite family that is relevant in nearly all applications (in this case the viscosity solution selects the distance function to* Γ*). For the interested reader, the reflected function* $v(x) = -u(x)$ *is also a viscosity solution, but of the negated equation* $-|u'(x)| = -1$*. Looking below to the definition of viscosity solution, we see that the sign of the equation is important.*

We now proceed to introduce the notion of viscosity solution, and review some properties of the state-constrained eikonal equation (29). Throughout this section we assume f is positive and Lipschitz continuous, $\Omega \subset \mathbb{R}^d$ is an open, bounded and connected domain, with a $C^{1,1}$ boundary $\partial\Omega$, and $\Gamma \subset \Omega$ is a closed set where we specify the homogeneous Dirichlet boundary conditions. In particular, we are not explicitly specifying boundary conditions on $\partial\Omega$, and instead we consider the *state constrained* problem (Capuzzo-Dolcetta and Lions,

1990). For the reader unfamiliar with PDE theory, we note that the assumption that $\partial\Omega$ is $C^{1,1}$ is equivalent to assuming there is a radius R such that at every boundary point $x \in \partial\Omega$, there exist balls of radius R touching x from inside and outside the domain (Lewicka and Peres, 2020). This is also equivalent to assuming the *reach*[4] of the boundary $\partial\Omega$, as a submanifold of $\mathbb{R}^d$ is lower bounded by R, and that the unit normal vector to the boundary is Lipschitz with constant $\frac{1}{R}$. Throughout this section we use the $C^{0,1}$ norm of a function, which is defined by

$$\|u\|_{C^{0,1}(\Omega)} = \|u\|_{C^0(\Omega)} + \mathrm{Lip}(u), \tag{31}$$

where $\|u\|_{C^0(\Omega)} = \max_{x\in\overline{\Omega}} |u(x)|$ and

$$\mathrm{Lip}(u) = \sup_{\substack{x,y\in\overline{\Omega}\\ x\neq y}} \frac{|u(x)-u(y)|}{|x-y|}.$$

We review the definition of viscosity solution for the state constrained problem here.

Definition 17. *We say that $u \in C(\overline{\Omega})$ is a* viscosity subsolution *of the state constrained equation* (29) *if $u \leq 0$ on Γ and if for each $x \in \Omega \setminus \Gamma$ and each $\varphi \in C^\infty(\mathbb{R}^d)$ such that $u - \varphi$ has a local maximum at x, we have*

$$|\nabla\varphi(x)| \leq f(x). \tag{32}$$

We say that $v \in C(\overline{\Omega})$ is a viscosity supersolution *of the state constrained equation* (29) *if $v \geq 0$ on Γ and if for each $x \in \overline{\Omega} \setminus \Gamma$ and each $\varphi \in C^\infty(\mathbb{R}^d)$ such that $v - \varphi$ has a local minimum at x, relative to $\overline{\Omega}$, we have*

$$|\nabla\varphi(x)| \geq f(x). \tag{33}$$

We say that u is a viscosity solution *of* (29) *if u is both a viscosity subsolution and a viscosity supersolution.*

Notice the key difference between the super and subsolution definitions is that in state constrained problems, we require the supersolution property to hold on the boundary $\partial\Omega$, but do not require the same in the subsolution property. We give a simple justification for this fact in the connection with the variational interpretation below. For a reference on viscosity solutions of Hamilton-Jacobi equations and connections to optimal control, we refer the reader to (Bardi et al., 1997), while as a reference for state constrained Hamilton-Jacobi equations, we refer to (Capuzzo-Dolcetta and Lions, 1990).

We quote below a comparison principle for state constrained Hamilton-Jacobi equations.

Theorem 18 (Capuzzo-Dolcetta and Lions, 1990)**.** *If u is a viscosity subsolution of* (29) *and v is a viscosity supersolution, then $u \leq v$ on $\overline{\Omega}$.*

It follows from Theorem 18 that solutions of (29) are unique. Existence can be obtained with the Perron method, or through the variational interpretation, which we discuss next.

4. The reach of a $\partial\Omega$ is the largest $R > 0$ such that any point x strictly within distance R of $\partial\Omega$ has a unique closest point in $\partial\Omega$.

3.1 Variational interpretation

The variational interpretation of (29) states that the solution of (29) is essentially a distance function on Ω to the set Γ, where distance is weighted by the positive function f. In particular, we first define the pairwise geodesic distance

$$d_f(x,y) := \inf\left\{\int_0^1 f(\gamma(t))|\gamma'(t)|\,dt \,:\, \gamma \in C^1([0,1];\overline{\Omega}), \gamma(0) = x, \text{ and } \gamma(1) = y\right\}. \tag{34}$$

The distance function $d_f(x,y)$ has been thoroughly studied in other works, we refer the reader to (Bardi et al., 1997; Calder, 2018a) for more details, and recall some relevant facts here. The function $d_f : \overline{\Omega} \times \overline{\Omega} \to \mathbb{R}$ is a metric, and in particular, it satisfies the triangle inequality

$$d_f(x,z) \leq d_f(x,y) + d_f(y,z).$$

We denote the distance function d_f with $f \equiv 1$ as

$$d_\Omega(x,y) = d_1(x,y).$$

The function $d_\Omega : \overline{\Omega} \times \overline{\Omega} \to \mathbb{R}$ is the *geodesic distance function* on Ω. Associated with the geodesic distance function, we define geodesic balls by

$$B_\Omega(x,r) = \{y \in \overline{\Omega} \,:\, d_\Omega(x,y) \leq r\}.$$

We will have to frequently utilize the geodesic distance $d_\Omega(x,y)$ in place of the Euclidean distance $|x-y|$, and we will need to compare the two distances. Since the boundary $\partial\Omega$ is $C^{1,1}$ and Ω is connected, there exists a constant $C > 0$, depending only on $\partial\Omega$ such that

$$|x-y| \leq d_\Omega(x,y) \leq |x-y| + C|x-y|^2 \quad \text{for all } x, y \in \overline{\Omega}. \tag{35}$$

We refer to Bungert et al. (2022, Proposition 5.1) for a proof of this fact. In fact, if the domain is *convex* then we have $d_\Omega(x,y) = |x-y|$, but we will not place such strong assumptions on the domain here. Associated with the geodesic distance, we also define the geodesic diameter

$$\text{diam}(\Omega) = \max_{x,y \in \overline{\Omega}} d_\Omega(x,y).$$

The geodesic diameter is finite, since Ω is connected and $\partial\Omega$ is $C^{1,1}$ (so that (35) holds).

Given the definition of the path distance function $d_f(x,y)$, we recall that the solution u of (29) is given by the variational representation formula

$$u(x) = \min_{y \in \Gamma} d_f(x,y). \tag{36}$$

Theorem 19. *The function u defined in* (36) *is the unique viscosity solution of* (29)*.*

The proof of Theorem 19 is standard in viscosity solution theory, and follows arguments in (Bardi et al., 1997) closely. We include a proof in Appendix Section B for the interested reader.

3.2 Lipschitz regularity

The variational interpretation of the eikonal equation gives a simple proof of Lipschitzness of the solution u.

Lemma 20. *Let $u \in C(\overline{\Omega})$ be the solution of (29). Then u is Lipschitz continuous and*

$$\mathrm{Lip}(u) \leq C\|f\|_{C^0(\Omega)},$$

where C depends only on $\mathrm{diam}(\Omega)$ and the $C^{1,1}$ bound on $\partial\Omega$.

Proof Since Ω is open and connected, we have $d_\Omega(x,y) < \infty$ for all $x, y \in \overline{\Omega}$. By (35), there exists $\widetilde{C} > 0$ such that

$$d_\Omega(x,y) \leq \widetilde{C}|x-y| \quad \text{for all } x, y \in \overline{\Omega} \text{ with } |x-y| \leq 1.$$

For $|x-y| \geq 1$ we have

$$d_\Omega(x,y) \leq \mathrm{diam}(\Omega) \leq \mathrm{diam}(\Omega)|x-y|.$$

Therefore

$$d_\Omega(x,y) \leq C|x-y| \quad \text{for all } x, y \in \overline{\Omega},$$

where $C = \max\{\widetilde{C}, \mathrm{diam}(\Omega)\}$.

Using Theorem 19 and the dynamic programming principle we have

$$u(y) \leq u(x) + d_f(x,y) \leq u(x) + \|f\|_{C^0(\Omega)} d_\Omega(x,y).$$

Swapping the roles of x and y yields

$$|u(x) - u(y)| \leq \|f\|_{C^0(\Omega)} d_\Omega(x,y) \leq C\|f\|_{C^0(\Omega)}|x-y|,$$

which completes the proof. ∎

3.3 Domain perturbations

In our discrete to continuum convergence theory in Section 4 below, we will need some results on the stability of the solution u of (29) under perturbations in the domain Ω. Let us define the signed distance function to the boundary $\partial\Omega$ by

$$d_{\partial\Omega}(x) = \begin{cases} \mathrm{dist}(x, \partial\Omega), & \text{if } x \in \overline{\Omega} \\ -\mathrm{dist}(x, \partial\Omega), & \text{otherwise.} \end{cases}$$

For $\delta \in \mathbb{R}$ we also define

$$\Omega_\delta = \{x \in \mathbb{R}^d : d_{\partial\Omega}(x) > \delta\} \quad \text{and} \quad \partial_\delta\Omega = \{x \in \mathbb{R}^d : |d_{\partial\Omega}(x)| \leq \delta\}. \tag{37}$$

We also recall the positive and negative part notation $\delta_+ = \max\{\delta, 0\}$ and $\delta_- = \min\{\delta, 0\}$.

Theorem 21. *For $\delta \in \mathbb{R}$ let $u_\delta \in C(\overline{\Omega_\delta})$ denote the viscosity solution of*

$$\begin{cases} |\nabla u_\delta| = f, & \text{in } \Omega_\delta \setminus \Gamma \\ u_\delta = 0, & \text{on } \Gamma, \end{cases} \tag{38}$$

and let $u = u_0$ be the viscosity solution of (29). There exists $C, c > 0$, depending only on $\partial\Omega$ and $\operatorname{dist}(\Gamma, \partial\Omega)$, such that whenever $|\delta| \leq c$ the following hold.

(i) $\operatorname{Lip}(u_\delta) \leq C\|f\|_{C^0(\Omega)}$, *and*

(ii) $\|u - u_\delta\|_{C^0(\Omega_{\delta_+})} \leq C f_{min}^{-1} \|f\|_{C^{0,1}(\Omega_{\delta_-})}\delta$, *where* $f_{min} = \min_{\Omega_{\delta_-}} f$.

Proof Since the boundary $\partial\Omega$ is $C^{1,1}$, the reach of $\partial\Omega$ is bounded below by a positive number $R > 0$ (in fact, $\frac{1}{R}$ is the Lipschitz constant of the unit normal vector to the boundary, we refer to (Lewicka and Peres, 2020) for details). Hence, within the tube $\partial_{\frac{R}{2}}\Omega$, the signed distance function $d_{\partial\Omega}$ is uniformly $C^{1,1}$. Hence, the perturbed boundaries $\partial\Omega_\delta$ are uniformly $C^{1,1}$ for $|\delta| \leq \frac{R}{4}$. Invoking Lemma 20 proves (i). We take $c \leq \frac{R}{2}$ smaller, if necessary, so that $\Gamma \subset \Omega_c$, and we assume $|\delta| \leq c$ for the rest of the proof.

We will prove the case of $\delta > 0$; the proof for $\delta < 0$ is very similar. It is clear that $u \leq u_\delta$ on Ω_δ, since there are more restrictions on the feasible paths in the variational interpretation of u_δ, compared to u. To prove the estimate in the other direction, that $u_\delta \leq u + C f_{min}^{-1}\|f\|_{C^{0,1}(\Omega)}\delta$, we use the comparison principle Theorem 18, with a suitable extension of u_δ to Ω. To do this, we define the cutoff function

$$\zeta(x) = \begin{cases} 1, & \text{if } 0 \leq d_{\partial\Omega}(x) \leq \frac{R}{4} \\ 2 - \frac{4}{R} d_{\partial\Omega}(x), & \text{if } \frac{R}{4} \leq d_{\partial\Omega}(x) \leq \frac{R}{2} \\ 0, & \text{if } d_{\partial\Omega}(x) \geq \frac{R}{2}. \end{cases} \tag{39}$$

The function ζ is a Lipschitz cutoff functions near the boundary $\partial\Omega$. Since $|\nabla d_{\partial\Omega}| = 1$ we have that $|\nabla\zeta| \leq \frac{4}{R} = C$, where C depends only on $\partial\Omega$, at all points of differentiability of ζ in $\partial_{\frac{R}{2}}\Omega$. We now define the extended function $w \in C(\overline{\Omega})$ by

$$w(x) = u_\delta(x + \delta\,\zeta(x)\nabla d_{\partial\Omega}(x)). \tag{40}$$

To shed light on the definition of w, we note that $\nabla d_{\partial\Omega}$ gives a natural extension of the unit inward normal vector ν from the boundary $\partial\Omega$ to the tube $\partial_R\Omega$. Indeed, $\nabla d_{\partial\Omega}$ agrees with the unit inward normal vector on the boundary $\partial\Omega$, and in fact, $\nabla d_{\partial\Omega}(x) = \nu(x_*)$, where $x_* \in \partial\Omega$ is the closest point to x from the boundary. Thus, we are simply stretching u_δ onto the larger domain Ω.

We first check that w is well-defined. If $x \in \Omega \setminus \Omega_{\frac{R}{4}}$, then $\zeta(x) = 1$ and so

$$d_{\partial\Omega}(x + \delta\,\zeta(x)\nabla d_{\partial\Omega}(x)) = d_{\partial\Omega}(x + \delta\nabla d_{\partial\Omega}(x)) = d_{\partial\Omega}(x) + \delta > \delta.$$

Hence $x + \delta\,\zeta(x)\nabla d_{\partial\Omega}(x) \in \Omega_\delta$ belongs to the domain of u_δ. If $x \in \Omega_{\frac{R}{4}}$, then

$$d_{\partial\Omega}(x + \delta\zeta(x)\nabla d_{\partial\Omega}(x)) = d_{\partial\Omega}(x) + \zeta(x)\delta \geq d_{\partial\Omega}(x) > \frac{R}{4} \geq \delta,$$

and we reach the same conclusion. This establishes that w is well-defined.

We will show that w is a viscosity subsolution of a similar equation, and then apply the comparison principle. To do this, we will use the fact that for a Hamiltonian that is convex in the gradient (i.e., the eikonal Hamiltonian $|\nabla u|$), Lipschitz continuous almost everywhere subsolutions are also viscosity subsolutions (the same is not true for supersolutions). This is a standard fact in viscosity solution theory, whose proof can be found in standard references (Bardi et al., 1997). Thus, we can work directly with ∇w at points of differentiability, instead of using the test function definition of viscosity solutions.

Let $x \in \Omega$, and assume that w and ζ are differentiable and x, and that $d_{\partial\Omega}$ is twice differentiable at x. Since $d_{\partial\Omega}$ is $C^{1,1}$, it is twice differentiable almost everywhere due to Radamacher's Theorem (Evans, 2010), so the set of such x has full measure. Then we compute

$$\nabla w(x) = [I + \delta(\zeta \nabla^2 d_{\partial\Omega}(x) + \nabla\zeta(x)) \otimes \nabla d_{\partial\Omega}(x)] \nabla u_\delta(x + \delta\zeta(x)\nabla d_{\partial\Omega}(x)).$$

Taking norms on both sides yields

$$\begin{aligned} |\nabla w(x)| &\leq \|I + \delta(\zeta \nabla^2 d_{\partial\Omega}(x) + \nabla\zeta(x)) \otimes \nabla d_{\partial\Omega}(x)\| |\nabla u(x + \delta\zeta(x)\nabla d_{\partial\Omega}(x))| \\ &\leq \left(1 + \delta\left(\|\nabla^2 d_{\partial\Omega}(x)\| + |\nabla\zeta(x)||\nabla d_{\partial\Omega}(x)|\right)\right) f(x + \delta\zeta(x)\nabla d_{\partial\Omega}(x)). \end{aligned}$$

Since $\nabla d_{\partial\Omega}$ is Lipschitz continuous in $\Omega_{\frac{R}{2}}$, we have a uniform bound on $\|\nabla^2 d_{\partial\Omega}\|$ at all points of differentiability. Thus, taking C larger, if necessary, we have

$$|\nabla w(x)| \leq (1 + C\delta) f(x + \delta\zeta(x)\nabla d_{\partial\Omega}(x)) \leq f(x) + C\|f\|_{C^{0,1}(\Omega)}\delta.$$

Set $v(x) = (1 + C f_{min}^{-1}\|f\|_{C^{0,1}(\Omega)}\delta)u(x)$. Then v is a viscosity solution of

$$|\nabla v(x)| \geq (1 + C f_{min}^{-1}\|f\|_{C^{0,1}(\Omega)}\delta) f(x) \geq f(x) + C\|f\|_{C^{0,1}(\Omega)}\delta.$$

By the comparison principle Theorem 18 we have $w \leq v$, and hence

$$u_\delta(x + \delta\,\zeta(x)\nabla d_{\partial\Omega}(x)) \leq u(x) + C f_{min}^{-1}\|f\|_{C^{0,1}(\Omega)}\delta$$

for all $x \in \Omega$. For $x \in \Omega_\delta$ we compute

$$u_\delta(x) \leq u_\delta(x + \delta\,\zeta(x)\nabla d_{\partial\Omega}(x)) + \mathrm{Lip}(u_\delta)\delta \leq u(x) + C f_{min}^{-1}\|f\|_{C^{0,1}(\Omega)}\delta + C\|f\|_{C^0(\Omega)}\delta,$$

which completes the proof. ∎

The domain perturbation result in Theorem 21 was the main objective in this section. Looking forward to Section 4, the domain perturbation allows us to handle the state-constrained boundary condition in the convergence proofs given in Theorems 22 and 23.

4. Discrete to continuum convergence

In this section we establish a continuum limit for the p-eikonal equation on a random geometric graph. In particular, we show that even though the p-eikonal equation does

not correspond to a graph distance function, its continuum limit does in fact recover the geodesic distance.

Let $x_1, x_2, \dots, x_n$ be a sequence of $i.i.d$ random variables on Ω with density ρ. As in Section 3 we assume that $\Omega \subset \mathbb{R}^d$ is open, bounded and connected with a $C^{1,1}$ boundary $\partial\Omega$. We assume the density ρ is Lipschitz continuous and bounded above and below by positive constants

$$\rho_{min} \leq \rho(x) \leq \rho_{max} \tag{41}$$

for all $x \in \Omega$. The vertices of the graph are denoted by

$$X := \{x_1, x_2, \dots, x_n\}. \tag{42}$$

To define the edges in a random geometric graph, we introduce a kernel $\eta : [0, \infty) \to [0, \infty)$, which is Lipschitz and nonincreasing and satisfies $\eta(0) > 0$ and $\eta(t) = 0$ for $t > 1$. For notational convenience, we also assume η has unit mass, so that

$$\int_{B(0,1)} \eta(|z|)\, dz = 1.$$

For $\varepsilon > 0$ define $\eta_\varepsilon(t) := \frac{1}{\varepsilon^d}\eta(\frac{t}{\varepsilon})$ and set $\sigma_p := \int_{\mathbb{R}^d} \eta_\varepsilon(|z|)|z_1|^p dz$. Note also that $\int_{B(0,\varepsilon)} \eta_\varepsilon(|z|)\, dz = 1$. The normalized weight w_{ij} between x_i and x_j is then given by

$$w_{ij} = \frac{\eta_\varepsilon(|x_i - x_j|)}{n\sigma_p \varepsilon^p}. \tag{43}$$

Letting $G_{n,\varepsilon}$ denote the graph with edge weights given in (43), the p-eikonal operator $\mathcal{A}_{G_{n,\varepsilon},p}$ is defined in (19), and is given by

$$\mathcal{A}_{G_{n,\varepsilon},p} u(x) := \frac{1}{n\sigma_p \varepsilon^p} \sum_{y \in X} \eta_\varepsilon\big(|x - y|\big)\big(u(x) - u(y)\big)_+^p. \tag{44}$$

For notational simplicity, we will write $\mathcal{A}_{n,\varepsilon} = \mathcal{A}_{G_{n,\varepsilon},p}$.

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

$$\begin{cases} \mathcal{A}_{n,\varepsilon} u(x) = f(x) & \text{if } x \in X \setminus \Gamma \\ u(x) = 0 & \text{if } x \in \Gamma, \end{cases} \tag{45}$$

where $\Gamma \subset X$ is a subset of the graph nodes where the homogeneous Dirichlet condition is set, and $f : \Omega \to \mathbb{R}$. We assume that f is Lipschitz continuous and that there exists $0 < f_{min} \leq f_{max}$ such that

$$0 < f_{min} \leq f(x) \leq f_{max} \quad \text{for all } x \in \Omega. \tag{46}$$

We also need to assume Γ is not too close to the topological boundary $\partial\Omega$. In particular, we assume

$$\operatorname{dist}(\Gamma, \partial\Omega) \geq R \tag{47}$$

where R is the reach of $\partial\Omega$. Other than this, we place no assumptions on Γ. We compare this graph equation to its continuum counterpart, the state-constrained eikonal equation

$$\begin{cases} \rho|\nabla u|^p = f & \text{in } \Omega \setminus \Gamma \\ u = 0 & \text{on } \Gamma. \end{cases} \tag{48}$$

It is important to point out that the set Γ is the same in (45) and (48), so the solution of (48) is in fact a random variable, depending on the locations of the points in Γ.

We recall from Section 3 that the solution of the state-constrained eikonal equation (48) is given by the geodesic distance function $u(x) = d_g(x, \Gamma)$ where $g := \rho^{-\frac{1}{p}} f^{\frac{1}{p}}$ (indeed, we simply rearrange (48) to read $|\nabla u| = g$).

Our main discrete to continuum convergence results are broken into two theorems, which are summarized below. In the theorem statements we write $u_{n,\varepsilon}$ for the solution of (45).

Theorem 22. *There exists $C, c > 0$ such that for ε sufficiently small and any $0 < \lambda \leq 1$ we have*

$$\mathbb{P}\left(\max_{x\in X}(d_g(x,\Gamma) - u_{n,\varepsilon}(x)) \leq C(\sqrt{\varepsilon} + \lambda)\right) \geq 1 - 2n\exp(-cn\varepsilon^d\lambda^2). \tag{49}$$

Theorem 23. *There exists $C, c > 0$ such that for ε sufficiently small and any $0 < \lambda \leq 1$ we have*

$$\mathbb{P}\left(\max_{x\in X}(u_{n,\varepsilon}(x) - d_g(x,\Gamma)) \leq C\left(\sqrt{\varepsilon} + \left(n\varepsilon^{p+d}\right)^{\frac{1}{p}} + \lambda\right)\right) \geq 1 - 3n^2\exp(-cn\varepsilon^d\lambda^2). \tag{50}$$

Remark 24. *The constants in both theorems depend on* $\operatorname{diam}(\Omega)$*, the $C^{1,1}$ bound on $\partial\Omega$ (or, equivalently, the reach of $\partial\Omega$), the kernel η (in particular $\eta(0)$), and the constants $r \in (0, 1]$ and $\mu > 0$ defined in Lemma 32), the dimension d, $\rho_{min}, \rho_{max}, \operatorname{Lip}(\rho), f_{min}, f_{max}, \operatorname{Lip}(f)$, and p. The dependence on p is uniform over compact sets, that is if $p \in [1, p_0]$, the constants depend only on p_0.*

Remark 25. *In order for the result of Theorem 23 to be non-vacuous, we require that*

$$n\varepsilon^{d+p} \ll 1. \tag{51}$$

For the probabilities in both Theorems 22 and 23 to be close to one, for arbitrarily small choices of $\lambda > 0$, we require $n\varepsilon^d \gg \log(n)$. Combining these two restrictions leads to the following restrictions on ε:

$$\left(\frac{\log(n)}{n}\right)^{\frac{1}{d}} \ll \varepsilon \ll \left(\frac{1}{n}\right)^{\frac{1}{p+d}}. \tag{52}$$

Since $p \geq 1$, there is always room between the upper and lower bounds to select a feasible ε. In general, we believe the upper bound is tight. Figure 6 shows the solution of the p-eikonal equation with $\Gamma = \{0\}$, giving a cone-like function, for different choices of p and ε. When the upper bound is violated, we see a spike forming at Γ, and the solution will fail to attain the boundary condition $u = 0$ on Γ in the continuum limit (note that this spike is utilized in our Lipschitz estimate in Section 4.2). We do expect, however, that the upper bound in

(52) *can be relaxed if we place more assumptions on the boundary nodes* Γ*, so that isolated points need not be considered. In particular, if* Γ *contains all points within distance* ε *of* $\partial\Omega$*, then the upper bound can be dropped using arguments from (Calder et al., 2020b).*

Finally, we note that there is some precedent for bandwidth restrictions like (52) *in the analysis of p-Laplacian semi-supervised learning in the same setting of arbitrarily low label rates. In (Slepčev and Thorpe, 2019) it was shown that p-Laplacian semi-supervised learning at low label rates requires the much more restrictive condition*

$$\left(\frac{\log(n)}{n}\right)^{\frac{1}{d}} \ll \varepsilon \ll \left(\frac{1}{n}\right)^{\frac{1}{p}},$$

which is only true when $p > d$*.*

Remark 26. *If we choose* $\lambda = \sqrt{\varepsilon}$ *in Theorems 22 and 23, then we obtain that the convergence rate*

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

holds with probability at least $1 - 5n^2\exp\left(-cn\varepsilon^{d+1}\right)$*. If we additionally choose* ε *so that* $\left(n\varepsilon^{p+d}\right)^{\frac{1}{p}} \leq \sqrt{\varepsilon}$*, that is, we make the restriction*

$$\varepsilon \leq \left(\frac{1}{n}\right)^{\frac{2}{p+2d}},$$

then the rate

$$\max_{x\in X} |u_{n,\varepsilon}(x) - d_g(x,\Gamma)| \leq C\sqrt{\varepsilon}$$

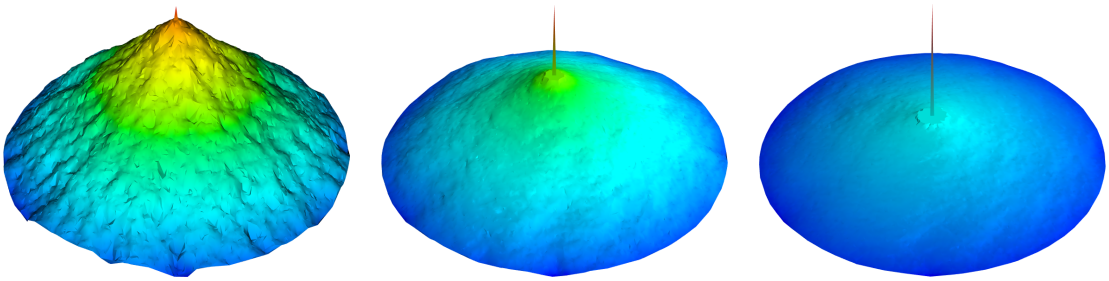
holds with the same probability. Without any further assumptions on the boundary set Γ*, we expect the* $\mathcal{O}(\sqrt{\varepsilon})$ *rate is optimal, in this general setting.*

Remark 27. *We expect that the results in this section will hold for other geometric graph constructions, with minor modifications to the statements. In particular, for real data it is very common to use k-nearest neighbor graphs, since they have better sparsity properties and are better adapted to the data. Since we use k-nearest neighbor graphs in all of our experiments with real data in Section 6, it is worth discussing briefly how the continuum limit results given in Theorem 22 and 23 would change.*

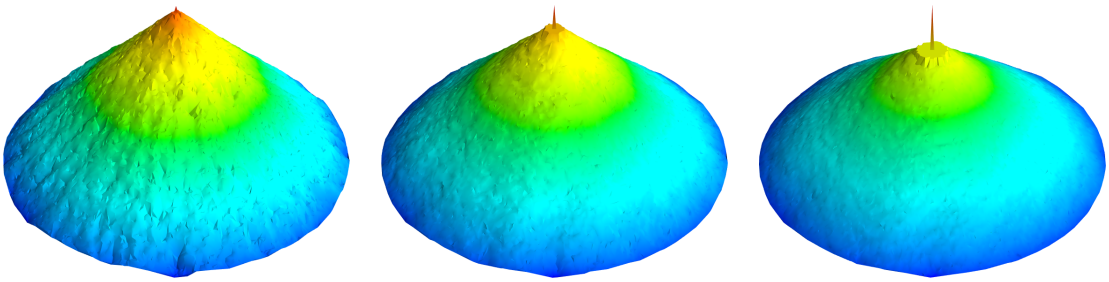
A k-nearest neighbor graph can be viewed as a random geometric graph with weights given by (43) *where the bandwidth* ε *is locally adapted to be the distance to the* k^{th} *nearest neighbor. Let* $\varepsilon_k(x)$ *denote the distance from* x *to its* k^{th} *nearest neighbor, and note that* $\varepsilon_k(x)$ *approximately satisfies* $n\varepsilon_k(x)^d\rho(x)\omega_d = k$*, where* $\omega_d = |B(0,1)|$ *is the measure of the unit ball.*[5] *Therefore*

$$\varepsilon_k(x) \approx \left(\frac{k}{n\rho(x)\omega_d}\right)^{\frac{1}{d}}. \tag{53}$$

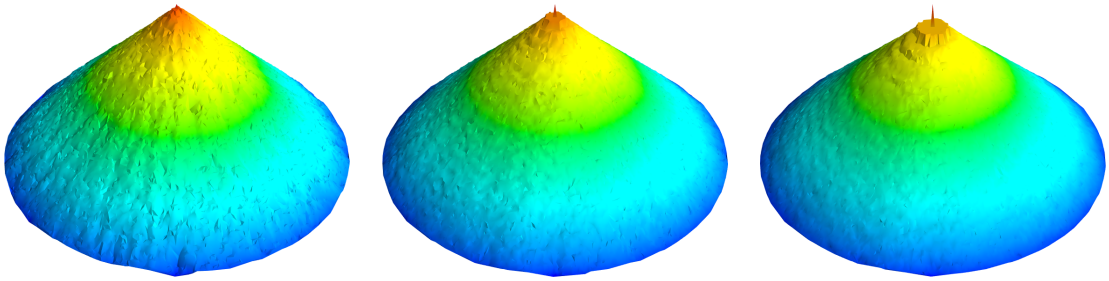
5. The quantity $\varepsilon_k(x)^d\rho(x)\omega_d$ approximates the probability mass of the ball $B(x,\varepsilon_k(x))$ when $\varepsilon_k(x)$ is small.



(a) $\varepsilon = 0.03, p = 1$ (b) $\varepsilon = 0.06, p = 1$ (c) $\varepsilon = 0.09, p = 1$



(d) $\varepsilon = 0.03, p = 2$ (e) $\varepsilon = 0.06, p = 2$ (f) $\varepsilon = 0.09, p = 2$



(g) $\varepsilon = 0.03, p = 4$ (h) $\varepsilon = 0.06, p = 4$ (i) $\varepsilon = 0.09, p = 4$

Figure 6: Simulations showing the solution of the p-eikonal equation with $\Gamma = \{0\}$ and $f \equiv 1$ (so we do not use density weighting) for different values of p and ε. The cones are inverted for a better viewing angle. The graph weights are given by (43) with $\eta(t) = \mathbb{1}_{[0,1]}(t)$. We see the size of the spike, which our theory shows is $n\varepsilon^{d+p}$, increases with ε and decreases with p, as expected. For reference, for $\varepsilon = 0.03$ each node in the graph has on average approximately 20 neighbors, while for $\varepsilon = 0.06$ and $\varepsilon = 0.09$ each node has on average 70 and 160 neighbors, respectively.

The approximation becomes better as $n \to \infty$, and so $\varepsilon_k \to 0$ (depending on the scaling of $k = k_n$ of course). Thus, if we were to normalize the p-eikonal equation on a k-nearest neighbor graph in the form

$$\frac{1}{n\sigma_p \varepsilon_k(x)^p} \sum_{y \in X} \eta_{\varepsilon_k(x)}(|x-y|)(u(x)-u(y))_+^p = f(x), \tag{54}$$

then the continuum limit would be the same eikonal equation (48)*. However, this type of normalization is not common, since one often does not have access to the distances* $\varepsilon_k(x)$*, and so a more common normalization would be to replace* $\varepsilon_k(x)^p$ *with* $\left(\frac{k}{n\omega_d}\right)^{\frac{p}{d}}$*, due to the approximation* (53)*. In this case, the p-eikonal operator on a k-nearest neighbor graph would be*

$$\mathcal{A}_{n,k}u(x) := \frac{1}{n\sigma_p}\left(\frac{n\omega_d}{k}\right)^{\frac{p}{d}} \sum_{y\in X} \eta_{\varepsilon_k(x)}(|x-y|)(u(x)-u(y))_+^p = f(x). \tag{55}$$

The only difference between (54) *and* (55)*, in the continuum limit, is the missing term* $\rho(x)^{\frac{p}{d}}$*. This suggests that the continuum limit of the equation* (55) *is given by*

$$\rho^{1-\frac{p}{d}}|\nabla u|^p = f.$$

The solution of this equation is the weighted geodesic distance $d_g(x,\Gamma)$ *where* $g = \rho^{\frac{1}{d}-\frac{1}{d}} f^{\frac{1}{p}}$*.*

4.1 Pointwise consistency

The first ingredient for a discrete to continuum limit result is pointwise consistency for the operator $\mathcal{A}_{n,\varepsilon}$. As usual, pointwise consistency passes through a nonlocal operator, which in this case has the form

$$\mathcal{A}_\varepsilon u(x) := \frac{1}{\sigma_p \varepsilon^p}\int_\Omega \eta_\varepsilon(|x-y|)(u(x)-u(y))_+^p \rho(y)dy, \tag{56}$$

for $u \in C^0(\Omega)$. Pointwise consistency is obtained in two steps, the first step (Lemma 28) passes from the discrete operator $\mathcal{A}_{n,\varepsilon}$ to the nonlocal counterpart $\mathcal{A}_\varepsilon$ via concentration of measure, while the second (Lemma 30) uses Taylor expansion to relate the nonlocal operator to the eikonal equation.

Lemma 28. *(Discrete to nonlocal) Let* $u : \overline{\Omega} \to \mathbb{R}$ *be Lipschitz continuous and* $n \geq 2$*. Then for any* $\lambda > 0$ *we have that*

$$\max_{x\in X} |\mathcal{A}_{n,\varepsilon}u(x) - \mathcal{A}_\varepsilon u(x)| \leq \eta(0)\rho_{max}\,\mathrm{Lip}(u)^p\lambda \tag{57}$$

holds with probability at least

$$1 - 2n\exp\left(\frac{-\eta(0)\sigma_p^2\rho_{max}n\varepsilon^d\lambda^2}{4\left(1+\frac{1}{3}\sigma_p\eta(0)\lambda\right)}\right). \tag{58}$$

Remark 29. *We note that to ensure the probability in* (58) *is close to* 1*, when* $\lambda > 0$ *can be arbitrarily small, we require that*

$$n\varepsilon^d \gg \log(n).$$

This is the same restriction required for graph connectivity in random geometric graphs (Penrose, 2003) (more correctly, the restriction for graph connectivity is $n\varepsilon^d \geq C\log(n)$ *for*

a large enough constant C). In contrast, pointwise consistency for graph Laplacians requires a more restrictive length scale restriction of the form $n\varepsilon^{d+2} \gg \log(n)$ (see, e.g., (Calder et al., 2020b)), which does not cover smaller bandwidths ε where the graph is still connected. The reason for this difference is that graph Laplacians are second order differential operators, and are normalized by an additional factor of ε to obtain meaningful continuum limits.

Proof [Proof of Lemma 28] Fix $x \in \Omega$ and let $Y_i := \eta_\varepsilon(|x - x_i|)(u(x) - u(x_i))_+^p$ so that

$$\mathcal{A}_{n,\varepsilon}u(x) = \frac{1}{\sigma_p \varepsilon^p} \frac{1}{n} \sum_{i=1}^n Y_i. \tag{59}$$

Then we compute

$$\mathbb{E}(Y_i) = \int_\Omega \eta_\varepsilon(|x - y|)(u(x) - u(y))_+^p \rho(y) dy \tag{60}$$

and for $\sigma^2 := \mathrm{Var}(Y_i)$

$$\begin{aligned} \sigma^2 \leq \mathbb{E}(Y_i^2) &= \int_{\Omega \cap B(x,\varepsilon)} \eta_\varepsilon^2(|x - y|)(u(x) - u(y))_+^{2p} \rho(y) dy \\ &\leq \rho_{\max} \operatorname{Lip}(u)^{2p} \varepsilon^{2p} \int_{B(x,\varepsilon)} \eta_\varepsilon^2(|x - y|) dy \\ &\leq \eta(0) \rho_{max} \operatorname{Lip}(u)^{2p} \varepsilon^{2p-d} \int_{B(x,\varepsilon)} \eta_\varepsilon(|x - y|) dy \\ &= \eta(0) \rho_{max} \operatorname{Lip}(u)^{2p} \varepsilon^{2p-d}. \end{aligned} \tag{61}$$

We also compute

$$|Y_i| = \eta_\varepsilon(|x - x_i|)|(u(x) - u(x_i))_+^p| \leq \eta_\varepsilon(|x - x_i|)|u(x) - u(x_i)|^p \leq \eta(0) \operatorname{Lip}(u)^p \varepsilon^{p-d}. \tag{62}$$

We now invoke Bernstein's inequality (see Appendix A) to obtain

$$\Big| \frac{1}{n} \sum_{i=1}^n Y_i - \int_\Omega \eta_\varepsilon(|x - y|)(u(x) - u(y))_+^p \rho(y) dy \Big| \leq t \tag{63}$$

holds with probability at least

$$1 - 2 \exp\left(\frac{-nt^2}{2\eta(0) \operatorname{Lip}(u)^p \varepsilon^{p-d} \left(\rho_{max} \operatorname{Lip}(u)^p \varepsilon^p + \frac{t}{3} \right)} \right).$$

Setting $t = \eta(0) \rho_{max} \sigma_p \operatorname{Lip}(u)^p \varepsilon^p \lambda$ for a new parameter $\lambda > 0$ we obtain

$$|\mathcal{A}_{n,\varepsilon}u(x) - \mathcal{A}_\varepsilon u(x)| \leq \eta(0) \rho_{max} \operatorname{Lip}(u)^p \lambda$$

with probability at least

$$1 - 2 \exp\left(\frac{-\eta(0) \sigma_p^2 \rho_{max} n \varepsilon^d \lambda^2}{2\left(1 + \frac{1}{3} \sigma_p \eta(0) \lambda\right)} \right).$$

The rest of proof is completed by conditioning on x_i and then applying a union bound. Indeed, conditioning on $x_i = x$, the other $n-1$ points form an *i.i.d.* sequence, and we note that

$$\mathcal{A}_{n,\varepsilon}u(x_i) := \frac{1}{n\sigma_p\varepsilon^p}\sum_{j\neq i}\eta_\varepsilon(|x-x_j|)(u(x_i)-u(x_j))_+^p$$

is exactly in the form considered above, except the sum is over $n-1$ *i.i.d.* random variables, instead of n. Thus, we can apply the argument above, replacing n with $n-1$, and then union bounding over $i = 1, \dots, n$. To simplify the probability we use the bound $n-1 \geq n/2$ for $n \geq 2$. ■

We now turn to comparing the nonlocal operator $\mathcal{A}_\varepsilon$ to its continuum counterpart $\rho|\nabla u|^p$.

Lemma 30. *(Nonlocal to local) There exists $C > 0$ such that for every $\varepsilon > 0$, $p \geq 1$ and $\varphi \in C^2(\mathbb{R}^d)$, the following hold.*

(i) *If* $\operatorname{dist}(x, \partial\Omega) \geq \varepsilon$ *then*

$$\left|\mathcal{A}_\varepsilon\varphi(x) - \rho(x)|\nabla\varphi(x)|^p\right| \leq CM\varepsilon,$$

where

$$M := \sigma_p^{-1}\|\rho\|_{C^{0,1}}\left(p(\operatorname{Lip}(\varphi) + \|\varphi\|_{C^2}\varepsilon)^{p-1}\|\varphi\|_{C^2} + 1\right).$$

(ii) *If* $\operatorname{dist}(x, \partial\Omega) < \varepsilon$ *then*

$$\mathcal{A}_\varepsilon\varphi(x) - \rho(x)|\nabla\varphi(x)|^p \leq CM\varepsilon.$$

For convenience, we recall that the $C^{0,1}$ norm is defined in (31).

Proof We first prove (i). Since $B(x,\varepsilon) \subset \Omega$, we make the change of variables $z := (y-x)/\varepsilon$ in the nonlocal operator (56) and obtain

$$\mathcal{A}_\varepsilon\varphi(x) = \frac{1}{\sigma_p\varepsilon^p}\int_{B(0,1)}\eta(|z|)(\varphi(x)-\varphi(x+\varepsilon z))_+^p\rho(x+\varepsilon z)dz. \tag{64}$$

Using the Taylor expansion

$$\rho(x+z\varepsilon) = \rho(x) + \mathcal{O}(\operatorname{Lip}(\rho)\varepsilon) \tag{65}$$

for $|z| \leq 1$ we have

$$\mathcal{A}_\varepsilon\varphi(x) = \frac{1}{\sigma_p\varepsilon^p}\int_{B(0,1)}\eta(|z|)(\varphi(x)-\varphi(x+\varepsilon z))_+^p\rho(x)dz + \mathcal{O}(\sigma_p^{-1}\operatorname{Lip}(\rho)^p\varepsilon)$$

We now use the Taylor expansion

$$\varphi(x) - \varphi(x+z\varepsilon) = \varepsilon z\cdot\nabla\varphi(x) + \mathcal{O}(\|\varphi\|_{C^2}\varepsilon^2) \tag{66}$$

to obtain

$$\mathcal{A}_\varepsilon\varphi(x) = \frac{1}{\sigma_p}\int_{B(0,1)} \eta(|z|)(z\cdot\nabla\varphi(x) + \mathcal{O}(\|\varphi\|_{C^2}\varepsilon))_+^p \rho(x)dz + \mathcal{O}(\sigma_p^{-1}\operatorname{Lip}(\rho)\varepsilon). \tag{67}$$

We make the change of variables $y = Az$ for an orthogonal matrix A such that $A\nabla\varphi(x) = |\nabla\varphi(x)|e_d$. Then we have that $z\cdot\nabla\varphi(x) = Az\cdot A\nabla\varphi(x) = |\nabla\varphi(x)|y_d$ and thereby

$$\mathcal{A}_\varepsilon\varphi(x) = \frac{1}{\sigma_p}\int_{B(0,1)} \eta(|y|)(|\nabla\varphi(x)|y_d + \mathcal{O}(\|\varphi\|_{C^2}\varepsilon))_+^p \rho(x)dy + \mathcal{O}(\sigma_p^{-1}\operatorname{Lip}(\rho)\varepsilon). \tag{68}$$

We now use the bound

$$|(a+t)_+^p - a_+^p| \le p(|a|+|t|)^{p-1}|t|,$$

for $a, t \in \mathbb{R}$ and $p \ge 1$, which follows from Taylor expansion, to obtain

$$(|\nabla\varphi(x)|y_d + \mathcal{O}(\|\varphi\|_{C^2}\varepsilon))_+^p = |\nabla\varphi(x)|^p(y_d)_+^p + \mathcal{O}\left(p(\operatorname{Lip}(\varphi) + \|\varphi\|_{C^2}\varepsilon)^{p-1}\|\varphi\|_{C^2}\varepsilon\right).$$

Substituting this above, we have

$$\mathcal{A}_\varepsilon\varphi(x) = \rho(x)|\nabla\varphi(x)|^p\frac{1}{\sigma_p}\int_{B(0,1)}\eta(|y|)(y_d)_+^p dy + M, \tag{69}$$

where

$$|M| \le C\sigma_p^{-1}\left(p(\operatorname{Lip}(\varphi) + \|\varphi\|_{C^2}\varepsilon)^{p-1}\|\varphi\|_{C^2}\rho_{max} + \operatorname{Lip}(\rho)\right)\varepsilon.$$

Applying the definition of σ_p and using the bound $\rho_{max}, \operatorname{Lip}(\rho) \le \|\rho\|_{C^{0,1}}$ completes the proof of (i).

The proof of (ii) proceeds in a similar way, except that on the first step, since $B(x,\varepsilon)\cap\partial\Omega \ne \varnothing$, the change of variables $z = (y-x)/\varepsilon$ yields

$$\mathcal{A}_\varepsilon\varphi(x) \le \frac{1}{\sigma_p\varepsilon^p}\int_{B(0,1)}\eta(|z|)(\varphi(x)-\varphi(x+\varepsilon z))_+^p\overline{\rho}(x+\varepsilon z)dz,$$

where $\overline{\rho}$ is any extension of ρ to $\mathbb{R}^d$ that preserves its Lipschitz constant. The proof then proceeds in the same way as (i). ∎

4.2 Lipschitz regularity

Since we allow for general closed Dirichlet boundary sets $\Gamma \subset \Omega$ in our discrete to continuum framework, our results require an *a priori* Lipschitz bound for the discrete solutions of (45). In this section we prove a Lipschitz estimate with the barrier method. The first ingredient is a lower bounds on the volume of the set $B_\Omega(x,r)\cap\Omega$.

Proposition 31. *For $r > 0$ sufficiently small, depending only on $\partial\Omega$, we have*

$$|B_\Omega(x,r)\cap\Omega| \ge c_d r^d \quad \text{for all } x \in \overline{\Omega},$$

where

$$c_d = \frac{\omega_{d-1}}{2^{\frac{3d+1}{2}}(d+1)},$$

and $\omega_d = |B(0,1)|$ denotes the volume of the unit ball in $\mathbb{R}^d$.

We postpone the proof of Proposition 31 to Appendix B, and proceed to define our barrier function for the Lipschitz estimate. For $y \in \mathbb{R}^d$ we define

$$\delta_y(x) = \begin{cases} 1, & \text{if } y = x \\ 0, & \text{otherwise.} \end{cases}$$

Our barrier function will be a *geodesic* cone with a jump (or spike) at the origin. In particular, we define

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

for $\beta > 0$ to be determined. We refer the reader to Figure 6 for an illustration of the barrier, for different size spikes (though the cones are inverted in the figure). The following lemma establishes the basic supersolution properties of our barrier function.

Lemma 32. *Let $y \in \overline{\Omega}$ and $\beta > 0$. Let $r \in (0, 1]$ and $\mu > 0$, such that $\eta(|t|) \geq \mu > 0$ for all $|t| \leq r$, and let c_d be the constant from Proposition 31. Then for ε sufficiently small, depending only on $\partial\Omega$, the following results hold:*

(i) *For $x \in \overline{\Omega} \setminus B(y, r\varepsilon)$ it holds that*

$$\mathbb{P}\left(\mathcal{A}_{n,\varepsilon} v_{\beta,y}(x) \geq \frac{c_d \mu r^{d+p}}{\sigma_p 2^{2d+p+1}}\right) \geq 1 - \exp\left(-\tfrac{c_d r^d}{2^{2d+3}} \rho_{min} n \varepsilon^d\right). \tag{70}$$

(ii) *For $x \in \overline{\Omega} \cap B(y, r\varepsilon) \setminus \{y\}$ we have*

$$\mathcal{A}_{n,\varepsilon} v_{\beta,y}(x) \geq \frac{\mu \beta^p}{\sigma_p n \varepsilon^{p+d}}. \tag{71}$$

Proof We will prove the two cases above separately.

(i) Assume $x \in \overline{\Omega} \setminus B(y, r\varepsilon)$ and let us define

$$D := \left\{ z \in B(x, r\varepsilon) \,:\, d_\Omega(x, y) - d_\Omega(y, z) \geq \frac{r\varepsilon}{2} \right\}.$$

Since $x \neq y$ we compute

$$\begin{aligned} \mathcal{A}_{n,\varepsilon} v_{\beta,y}(x) =& \frac{1}{n\sigma_p \varepsilon^p} \sum_{z \in X} \eta_\varepsilon(|x - z|)(\beta + d_\Omega(x, y) - \beta(1 - \delta_y(z)) - d_\Omega(y, z))_+^p \\ &\geq \frac{\mu}{n\sigma_p \varepsilon^{p+d}} \sum_{z \in X \cap B(x, r\varepsilon)} (d_\Omega(x, y) - d_\Omega(y, z))_+^p \\ &\geq \frac{\mu}{n\sigma_p \varepsilon^{p+d}} \sum_{z \in X \cap D} \left(\frac{r\varepsilon}{2}\right)_+^p \\ &= \frac{\mu r^p}{2^p n \sigma_p \varepsilon^d} \#(X \cap D). \end{aligned} \tag{72}$$

To bound the number of points in $D \cap X$, we use the Chernoff bound (see Appendix A), which produces the lower bound

$$\mathcal{A}_{n,\varepsilon} v_{\beta,y}(x) \geq \frac{\mu r^p}{2^{p+1} \sigma_p \varepsilon^d} |D \cap \Omega| \tag{73}$$

with probability at least $1 - \exp\left(-\frac{1}{8}\rho_{min}|D \cap \Omega|n\right)$.

We need to lower bound $|D \cap \Omega|$ to complete the proof. There exists $z_* \in \partial B(x, \frac{3r\varepsilon}{4})$ so that

$$d_\Omega(x, y) = d_\Omega(x, z_*) + d_\Omega(z_*, y).$$

Since $d_\Omega(x, z_*) \geq |x - z_*| = \frac{3r\varepsilon}{4}$ this becomes

$$d_\Omega(x, y) - d_\Omega(y, z_*) \geq \frac{3r\varepsilon}{4}.$$

It follows that $B_\Omega(z_*, \frac{r\varepsilon}{4}) \subset D$. Indeed, if $d_\Omega(z, z_*) \leq \frac{r\varepsilon}{4}$ then by the triangle inequality we have

$$d_\Omega(x, y) - d_\Omega(y, z) \geq d_\Omega(x, y) - d_\Omega(y, z_*) - d_\Omega(z, z_*) \geq \frac{3r\varepsilon}{4} - \frac{r\varepsilon}{4} = \frac{r\varepsilon}{2}.$$

Invoking Proposition 31 we have

$$|D \cap \Omega| \geq |B_\Omega(z_*, \tfrac{r\varepsilon}{4}) \cap \Omega| \geq c_d \left(\frac{r\varepsilon}{4}\right)^d,$$

for ε sufficiently small. Combining this with (73) completes the proof of (i).

(ii) Let $x \in \overline{\Omega} \cap B(y, r\varepsilon) \setminus \{y\}$, and compute

$$\begin{aligned} \mathcal{A}_{n,\varepsilon} v_{\beta,y}(x) &\geq \frac{\eta_\varepsilon(|x-y|)}{\sigma_p n \varepsilon^p} (v_{\beta,y}(x) - v_{\beta,y}(y))_+^p \\ &\geq \frac{\mu}{\sigma_p n \varepsilon^{p+d}} (\beta + d_\Omega(x, y) - d_\Omega(y, y))_+^p \\ &\geq \frac{\mu \beta^p}{\sigma_p n \varepsilon^{p+d}}, \end{aligned} \tag{74}$$

which completes the proof. ∎

We are now equipped to prove global Lipschitzness for the p-eikonal equation. The proof is based on the barrier method, using the barrier studied in Lemma 32.

Theorem 33. *Let u be the solution of (45). Let c_d, r, and μ be as defined as in Lemma 32. Define*

$$\gamma_p = \left(\frac{c_d r^{d+p}}{2^{2d+p+1}}\right)^{\frac{1}{p}} \quad \text{and} \quad c_p = \left(\frac{\sigma_p}{\mu}\right)^{\frac{1}{p}}.$$

Then it holds with probability at least $1 - n^2 \exp\left(-\frac{c_d r^d}{2^{2d+3}} \rho_{min} n \varepsilon^d\right)$ that

$$|u(x) - u(y)| \leq c_p \gamma_p^{-1} \max_X f^{\frac{1}{p}} \, d_\Omega(x, y) + \gamma_p \left(n\varepsilon^{p+d}\right)^{\frac{1}{p}}, \quad \text{for all } x, y \in X. \tag{75}$$

Proof We choose β in Lemma 32 to satisfy

$$\beta^p = \frac{c_d r^{d+p}}{2^{2d+p+1}} n \varepsilon^{p+d} = \gamma_p^p n \varepsilon^{p+d}, \tag{76}$$

and we set $v_y = v_{\beta,y}$. Then by Lemma 32 and a union bound, we have that

$$\mathcal{A}_{n,\varepsilon} v_y(x) \geq \frac{\mu \gamma_p^p}{\sigma_p} \quad \text{for all } x, y \in X, x \neq y, \tag{77}$$

holds with probability at least $1 - n^2 \exp\left(-\frac{c_d r^d}{2^{2d+3}} \rho_{min} n \varepsilon^d\right)$ (note that the n^2 comes from the union bound over the over the pairs $(x, y) \in X^2$ with $x \neq y$, for which there are $n^2 - n \leq n^2$ events). For the rest of the proof we assume this event holds.

Let us define

$$C = \left(\frac{\sigma_p}{\mu \gamma_p^p}\right)^{\frac{1}{p}} \max_X f^{\frac{1}{p}}.$$

Then since $\mathcal{A}_{n,\varepsilon}$ is p-homogeneous we have

$$\mathcal{A}_{n,\varepsilon}(C v_y)(x) = C^p \mathcal{A}_{n,\varepsilon} v_y(x) \geq \max_X f \geq \mathcal{A}_{n,\varepsilon} u(x),$$

for all $x, y \in X$ with $x \neq y$. Therefore, $C v_y$ is a supersolution, relative to the function $w(x) := u(x) - u(y)$ on the set $X \setminus (\Gamma \cup \{y\})$. Furthermore, $w(y) = u(y) - u(y) = 0 \leq C v_y(y)$ and for $x \in \Gamma$ we have $w(x) = u(x) - u(y) \leq 0 - u(y) \leq 0 \leq v_y(x)$. Thus, by the comparison principle (Lemma 6) we have that $u(x) - u(y) \leq C v_y(x)$ for all $x, y \in X$ with $x \neq y$, which becomes

$$u(x) - u(y) \leq \left(\frac{\sigma_p}{\mu \gamma_p^p}\right)^{\frac{1}{p}} \max_X f^{\frac{1}{p}}\, d_\Omega(x, y) + \beta.$$

Substituting the definition of β, and reversing the role of x and y to get an absolute value bound, completes the proof. ■

Remark 34. *Similar to Lemma 20, we can use the bound $d_\Omega(x, y) \leq C|x - y|$ to obtain that the solution u of (45) satisfies*

$$|u(x) - u(y)| \leq C\left(|x - y| + \left(n\varepsilon^{p+d}\right)^{\frac{1}{p}}\right),$$

with probability at least $1 - n^2 \exp\left(-cn\varepsilon^d\right)$, where C and c are constants whose precise values are given in Theorem 33.

4.3 Discrete to continuum convergence

We now proceed to prove our main discrete to continuum convergence results. The results are split into two theorems. Throughout the proof of Theorems 22 and 23, we use the convention that $0 \leq c \leq 1$ and $C \geq 1$ denote arbitrary constants, whose value can change from line to line, to reduce the notational burden.

Proof [Proof of Theorem 22] For $0 < \delta \leq c$, where $c > 0$ is given in Theorem 21, let u_δ denote the viscosity solution of (38) over the perturbed domain $\Omega_{-\delta} \setminus \Gamma$, defined in Theorem 21, except with g in place of f on the right hand side. For $0 < \theta < 1$ and $1 \leq \alpha \leq \varepsilon^{-1}$ we define the auxiliary function

$$\Phi(x, y) := (1 - \theta) u_\delta(x) - u_{n,\varepsilon}(y) - \frac{\alpha}{2}|x - y|^2, \qquad (x, y) \in \overline{\Omega_{-\delta}} \times X.$$

Let $(x_\alpha, y_\alpha) \in \overline{\Omega_{-\delta}} \times X$ be a point at which Φ is maximized over $\overline{\Omega_{-\delta}} \times X$. To see why the auxiliary function is useful, we first note that the equality

$$(1-\theta)u_\delta(x) - u_{n,\varepsilon}(x) = \Phi(x,x)$$

implies that

$$\max_X \left((1-\theta)u_\delta - u_{n,\varepsilon}\right) \leq \max_{x\in X} \Phi(x,x) \leq \Phi(x_\alpha, y_\alpha).$$

We also have

$$\max_X (u_\delta - u_{n,\varepsilon}) \leq \max_X \left((1-\theta)u_\delta - u_{n,\varepsilon}\right) + C\theta,$$

and by Theorem 21 (ii) we have $|u - u_\delta| \leq C\delta$, with C depending on f and ρ, where $u(x) = u_0(x) = d_g(x,\Gamma)$ and $g = \rho^{-\frac{1}{p}} f^{\frac{1}{p}}$. Therefore, we obtain the bound

$$\max_X (u - u_{n,\varepsilon}) \leq \Phi(x_\alpha, y_\alpha) + C(\theta + \delta). \tag{78}$$

Thus, we will obtain an error estimate on $u - u_{n,\varepsilon}$ by estimating $\Phi(x_\alpha, y_\alpha)$, while choosing the parameters θ and δ as small as possible, and optimizing over α.

Since $\Phi(x_\alpha, y_\alpha) \geq \Phi(y_\alpha, y_\alpha)$, we have

$$(1-\theta)u_\delta(x_\alpha) - u_{n,\varepsilon}(y_\alpha) - \frac{\alpha}{2}|x_\alpha - y_\alpha|^2 \geq (1-\theta)u_\delta(y_\alpha) - u_{n,\varepsilon}(y_\alpha). \tag{79}$$

By Theorem 21 (i), u_δ is Lipschitz continuous, and so

$$\frac{\alpha}{2}|x_\alpha - y_\alpha|^2 \leq (1-\theta)(u_\delta(x_\alpha) - u_\delta(y_\alpha)) \leq C|x_\alpha - y_\alpha|. \tag{80}$$

Hence we have the bound

$$|x_\alpha - y_\alpha| \leq C\alpha^{-1}. \tag{81}$$

Thus, for $\alpha > C\delta^{-1}$, we have $|x_\alpha - y_\alpha| < \delta$ and so $x_\alpha \in \Omega_{-\delta}$, since $y_\alpha \in \Omega$. We assume $\alpha > C\delta^{-1}$ throughout the rest of the proof.

We now have several cases to consider.

(i) If $y_\alpha \in \Gamma$, then $u_{n,\varepsilon}(y_\alpha) = 0 = u_\delta(y_\alpha)$ and so

$$u_\delta(x_\alpha) - u_{n,\varepsilon}(y_\alpha) = u_\delta(x_\alpha) - u_\delta(y_\alpha) \leq C|x_\alpha - y_\alpha| \leq C\alpha^{-1}. \tag{82}$$

Therefore

$$\Phi(x_\alpha, y_\alpha) \leq u_\delta(x_\alpha) - u_{n,\varepsilon}(y_\alpha) \leq C\alpha^{-1}.$$

(ii) If $x_\alpha \in \Gamma$, then $u_\delta(x_\alpha) = 0 = u_{n,\varepsilon}(x_\alpha)$ and thus

$$u_\delta(x_\alpha) - u_{n,\varepsilon}(y_\alpha) = u_{n,\varepsilon}(x_\alpha) - u_{n,\varepsilon}(y_\alpha) \leq 0, \tag{83}$$

since $u_{n,\varepsilon} \geq 0$. In this case we have $\Phi(x_\alpha, y_\alpha) \leq 0$.

(iii) We now consider the case of interior maxima; in particular, that $x_\alpha \in \Omega_{-\delta} \setminus \Gamma$ and $y_\alpha \in X \setminus \Gamma$. Our plan is to choose the parameter θ so that interior maxima are impossible,

and so this case need not be considered when estimating $\Phi(x_\alpha, y_\alpha)$. We first note that the map

$$x \mapsto u_\delta(x) - \frac{\alpha}{2}(1-\theta)^{-1}|x - y_\alpha|^2$$

attains its maximum at x_α over the open set $\Omega_{-\delta}$. Using $\varphi(x) = \frac{\alpha}{2}(1-\theta)^{-1}|x-y_\alpha|^2$ as a test function for the definition of viscosity subsolution for u_δ, we have

$$|p_\alpha| \leq (1-\theta)g(x_\alpha), \tag{84}$$

where $p_\alpha = \alpha(x_\alpha - y_\alpha)$. Likewise, the map $y \mapsto u_{n,\varepsilon}(y) + \frac{\alpha}{2}|x_\alpha - y|^2$ attains its minimum at $y_\alpha \in X$ over the point cloud X. Setting $\psi(y) := -\frac{\alpha}{2}|x_\alpha - y|^2$, we see that the inequality

$$u_{n,\varepsilon}(y_\alpha) - u_{n,\varepsilon}(y) \leq \psi(y_\alpha) - \psi(y)$$

holds for all $y \in X$. It follows that

$$f(y_\alpha) = \mathcal{A}_{n,\varepsilon} u_{n,\varepsilon}(y_\alpha) \leq \mathcal{A}_{n,\varepsilon}\psi(y_\alpha).$$

Using pointwise consistency (Lemmas 28 and 30), and noting that $\mathrm{Lip}(\psi) \leq C$, $\|\psi\|_{C^2} \leq C\alpha$, and $\nabla\psi(y_\alpha) = p_\alpha$ we obtain that

$$f(y_\alpha) \leq \rho(y_\alpha)|p_\alpha|^p + C(\alpha\varepsilon + \lambda),$$

holds for any $0 < \lambda \leq 1$ with probability at least $1 - 2n\exp\left(-cn\varepsilon^d\lambda^2\right)$, where C depends on p, σ_p, $\|\rho\|_{C^{0,1}}$, $\eta(0)$, and ρ_{max}, and c depends on $\eta(0)$, σ_p, and ρ_{max}. Dividing by ρ on both sides, and combining with (84) yields

$$g(y_\alpha)^p \leq (1-\theta)^p g(x_\alpha)^p + C(\alpha\varepsilon + \lambda).$$

Since $(1-\theta)^p \leq 1-\theta$, and g^p is Lipschitz, we can rearrange this and use (81) to obtain

$$\theta g(x_\alpha) - C(\alpha\varepsilon + \lambda) \leq g(x_\alpha)^p - g(y_\alpha)^p \leq C|x_\alpha - y_\alpha| \leq C\alpha^{-1}.$$

Since g is bounded below by a positive constant, this yields

$$\theta \leq C(\alpha^{-1} + \alpha\varepsilon + \lambda).$$

Hence, we set

$$\theta = (C+1)(\alpha^{-1} + \alpha\varepsilon + \lambda),$$

so that case (iii) cannot hold.

The proof is completed by noting that cases (i) and (ii) yield $\Phi(x_\alpha, y_\alpha) \leq C\alpha^{-1}$, and so (78) yields

$$\max_X (u - u_{n,\varepsilon}) \leq C(\alpha^{-1} + \alpha\varepsilon + \delta + \lambda).$$

Optimizing over α yields $\alpha = \frac{1}{\sqrt{\varepsilon}}$. We also made the restriction $\alpha \geq C\delta^{-1}$ earlier, so we choose $\delta \geq C\sqrt{\varepsilon}$. Recalling that $u(x) = d_g(x, \Gamma)$ (see Theorem 19), the proof is complete. ■

Below we give the proof of Theorem 23. The main difference with the proof of Theorem 22 is that we now need to use the discrete Lipschitz estimate proved in Theorem 33. This introduces the additional error term $n\varepsilon^{d+p}$ and modifies the probability with which the convergence rate holds.

Proof [Proof of Theorem 23] The start of the proof is similar to Theorem 22. For $0 < \delta \leq c$, where $c > 0$ is given in Theorem 21, let u_δ denote the viscosity solution of (38) over the perturbed domain $\Omega_\delta \setminus \Gamma$, defined in Theorem 21, except with g in place of f on the right hand side. For $0 < \theta \leq 1$ and $1 \leq \alpha \leq \varepsilon^{-1}$ we define the auxiliary function

$$\Phi(x, y) := u_{n,\varepsilon}(x) - (1+\theta)u_\delta(y) - \frac{\alpha}{2}|x-y|^2, \qquad (x, y) \in X \times \overline{\Omega_\delta}.$$

Let $(x_\alpha, y_\alpha) \in X \times \overline{\Omega_\delta}$ be a point at which Φ is maximized over $X \times \overline{\Omega_\delta}$. As in the proof of Theorem 22 we have

$$\max_{X \cap \overline{\Omega_\delta}} (u_{n,\varepsilon} - u_\delta) \leq \Phi(x_\alpha, y_\alpha) + C\theta,$$

where $u(x) = u_0(x) = d_g(x, \Gamma)$. By the Lipschitzness of $u_{n,\varepsilon}$ (see Theorem 33 and Remark 34) and that of u_δ (see Theorem 21 (i)), this yields

$$\max_X (u_{n,\varepsilon} - u) \leq \Phi(x_\alpha, y_\alpha) + C\left(\theta + \delta + \left(n\varepsilon^{p+d}\right)^{\frac{1}{p}}\right), \tag{85}$$

with probability at least $1 - n^2 \exp(-cn\varepsilon^d)$. As in the proof of Theorem 22, the proof proceeds by estimating $\Phi(x_\alpha, y_\alpha)$, while choosing the parameters θ, δ and α appropriately.

Since $\Phi(x_\alpha, y_\alpha) \geq \Phi(x_\alpha, x_\alpha)$, we have

$$u_{n,\varepsilon}(x_\alpha) - (1+\theta)u_\delta(y_\alpha) - \frac{\alpha}{2}|x_\alpha - y_\alpha|^2 \geq u_{n,\varepsilon}(x_\alpha) - (1+\theta)u_\delta(x_\alpha). \tag{86}$$

Since u_δ is Lipschitz continuous we have

$$\frac{\alpha}{2}|x_\alpha - y_\alpha|^2 \leq (1+\theta)(u_\delta(x_\alpha) - u_\delta(y_\alpha)) \leq C|x_\alpha - y_\alpha|. \tag{87}$$

Hence we obtain the same bound $|x_\alpha - y_\alpha| \leq C\alpha^{-1}$ as in (81) from Theorem 22. We now make the restriction $\delta \geq 2\varepsilon$, and $C\alpha^{-1} \leq \delta$ so that $|x_\alpha - y_\alpha| \leq \varepsilon$. Since $y_\alpha \in \overline{\Omega_\delta}$, this ensures that

$$\operatorname{dist}(x_\alpha, \partial\Omega) \geq \operatorname{dist}(y_\alpha, \partial\Omega) - |x_\alpha - y_\alpha| \geq \delta - \varepsilon \geq \varepsilon.$$

Therefore $B(x_\alpha, \varepsilon) \subset \Omega$, which will allow us to utilize the pointwise consistency results (Lemmas 28 and 30) later on in the proof.

We again have several cases to consider.

(i) If $y_\alpha \in \Gamma$, then $u_\delta(y_\alpha) = 0 = u_{n,\varepsilon}(y_\alpha)$ and so by the Lipschitz continuity of $u_{n,\varepsilon}$ (see Remark 34) we have

$$\begin{aligned} u_{n,\varepsilon}(x_\alpha) - u_\delta(y_\alpha) &= u_{n,\varepsilon}(x_\alpha) - u_{n,\varepsilon}(y_\alpha) \\ &\leq C\left(|x_\alpha - y_\alpha| + \left(n\varepsilon^{p+d}\right)^{\frac{1}{p}}\right) \\ &\leq C\left(\alpha^{-1} + \left(n\varepsilon^{p+d}\right)^{\frac{1}{p}}\right). \end{aligned}$$

Therefore

$$\Phi(x_\alpha, y_\alpha) \leq u_{n,\varepsilon}(x_\alpha) - u_\delta(y_\alpha) \leq C\left(\alpha^{-1} + \left(n\varepsilon^{p+d}\right)^{\frac{1}{p}}\right).$$

(ii) If $x_\alpha \in \Gamma$, then $u_{n,\varepsilon}(x_\alpha) = 0 = u_\delta(x_\alpha)$ and thus

$$u_{n,\varepsilon}(x_\alpha) - u_\delta(y_\alpha) = u_\delta(x_\alpha) - u_\delta(y_\alpha) \leq 0, \tag{88}$$

since $u_\delta \geq 0$.

(iii) We now consider the case of $x_\alpha \in X \setminus \Gamma$ and $y_\alpha \in \overline{\Omega_\delta} \setminus \Gamma$, and we again show that θ can be chosen to rule out this case. We first note that the map

$$y \mapsto u_\delta(y) + \frac{\alpha}{2}(1+\theta)^{-1}|x_\alpha - y|^2$$

attains its minimum at $y_\alpha \in \overline{\Omega_\delta}$ relative to the closed set $\overline{\Omega_\delta}$. Using $\varphi(x) = -\frac{\alpha}{2}(1+\theta)^{-1}|x_\alpha - y|^2$ as a test function for the definition of viscosity supersolution for u_δ, and recalling from Definition 17 that the supersolution condition holds even on the boundary $\partial\Omega_\delta$, we have

$$|p_\alpha| \geq (1+\theta)g(y_\alpha), \tag{89}$$

where $p_\alpha = \alpha(x_\alpha - y_\alpha)$. Likewise, the map $x \mapsto u_{n,\varepsilon}(x) - \frac{\alpha}{2}|x - y_\alpha|^2$ attains its maximum at $x_\alpha \in X$ over the point cloud X. Setting $\psi(x) := \frac{\alpha}{2}|x - y_\alpha|^2$, we see that the inequality

$$u_{n,\varepsilon}(x_\alpha) - u_{n,\varepsilon}(x) \geq \psi(x_\alpha) - \psi(x)$$

holds for all $x \in X$. It follows that

$$f(x_\alpha) = \mathcal{A}_{n,\varepsilon}u_{n,\varepsilon}(x_\alpha) \geq \mathcal{A}_{n,\varepsilon}\psi(x_\alpha).$$

Since $B(x_\alpha, \varepsilon) \subset \Omega$, we can use pointwise consistency (Lemmas 28 and 30) to obtain

$$f(x_\alpha) \geq \rho(x_\alpha)|p_\alpha|^p - C(\alpha\varepsilon + \lambda),$$

holds for any $0 < \lambda \leq 1$ with probability at least $1 - 2n\exp\left(-cn\varepsilon^d\lambda^2\right)$, where C depends on p, σ_p, $\|\rho\|_{C^{0,1}}$, $\eta(0)$, and ρ_{max}, and c depends on $\eta(0)$, σ_p, and ρ_{max}. Dividing by ρ on both sides, and combining with (89) yields

$$g(x_\alpha)^p + C(\alpha\varepsilon + \lambda) \geq (1+\theta)^p g(y_\alpha)^p.$$

Since $(1+\theta)^p \geq 1 + \theta$, and g^p is Lipschitz, we can rearrange this and use (81) to obtain

$$\theta g(y_\alpha)^p - C(\alpha\varepsilon + \lambda) \leq g(x_\alpha)^p - g(y_\alpha)^p \leq C|x_\alpha - y_\alpha| \leq C\alpha^{-1}.$$

Since g is bounded below by a positive constant, this yields

$$\theta \leq C(\alpha^{-1} + \alpha\varepsilon + \lambda).$$

Hence, we set

$$\theta = (C+1)(\alpha^{-1} + \alpha\varepsilon + \lambda),$$

so that case (iii) cannot hold.

The proof is completed by combining cases (i) and (ii) with (85) to obtain

$$\max_X(u_{n,\varepsilon} - u) \leq C\left(\alpha^{-1} + \alpha\varepsilon + \delta + \left(n\varepsilon^{p+d}\right)^{\frac{1}{p}} + \lambda\right).$$

Optimizing over α yields $\alpha = \frac{1}{\sqrt{\varepsilon}}$. We also made the restrictions $\delta \geq 2\varepsilon$ and $\delta \geq C\alpha^{-1} = C\sqrt{\varepsilon}$. Thus, we can again choose $\delta = C\sqrt{\varepsilon}$ to satisfy these conditions, which completes the proof. ■

5. Continuum analysis

Given the discrete to continuum convergence results from Section 4, which show that the solution of the p-eikonal equation converges to a density weighted geodesic distance, we now proceed to study the asymptotic consistency of the p-eikonal equation for both data depth and semi-supervised learning. Throughout this section we let Ω be an open, connected domain, and denote by ρ the density function on Ω.

5.1 Background on medians and data depth

Before proceeding with our analysis, we give a background on medians in one dimension, and their extensions to higher dimensions, as well as various notions of data depth. In one dimension, the *median* of data points $x_1, x_2, \ldots, x_n \in \mathbb{R}$ is any point x minimizing

$$\sum_{i=1}^{n} |x - x_i|. \tag{90}$$

If n is odd, then $x = x_i$, where x_i is the middle point after sorting the datapoints from smallest to largest. If n is even, then any x between the two middle points is a valid median. We obtain a population level version of the median by assuming the x_i are *i.i.d.* with density ρ and taking the expectation of (90) to obtain the problem

$$\min_{x\in\mathbb{R}} \int_{-\infty}^{\infty} |x - y|\rho(y)\, dy.$$

The population level median is thus any x satisfying

$$\int_x^{\infty} \rho\, dy = \int_{-\infty}^{x} \rho\, dy = \frac{1}{2}.$$

Writing the notion of median in this way gives a natural extension to higher dimensions, $x, x_i \in \mathbb{R}^d$, by simply replacing the absolute values in (90) by the Euclidean norm. In this case, the notion of median is called the *geometric median* (see, e.g., (Minsker, 2015)). In fact, in even more general settings, we can replace the norm by any metric d, yielding a generalized geometric median, or *barycenter* problem

$$\min_x \sum_{i=1}^{n} d(x, x_i). \tag{91}$$

Here, we may take (X, d) to be a metric space and $x, x_i \in X$, but in general, we may consider (91) even when d is not a true metric. This approach is taken, for example, in the manifold setting by Fletcher et al. (2009). Our approach to medians follows this generalized geometric median approach, where the density weighting is involved in the metric d. We note that the problem (91) has the population level analog

$$\min_x \int_{\mathbb{R}^n} d(x, y)\rho(y)\, dy, \tag{92}$$

when the data points x_i are sampled from a probability distribution ρ.

Our approach to data depth is to first compute the median and then take the depth as the distance to the median. There are many other approaches to data depth that proceed in the opposite direction, first defining data depth and then taking the median to be the deepest point. For example, in one dimension, we can order the points from smallest to largest, and define the depth as the fewest number of points to the right or left. The population level version of this in one dimension is

$$\text{depth}(x) = \min\left\{\int_x^\infty \rho\, dy, \int_{-\infty}^x \rho\, dy\right\}.$$

Any point with maximum depth is clearly the (population level) median. A natural extension of this to higher dimensions is the Tukey depth (Tukey, 1975), which is given by

$$\text{Tukey Depth(x)} = \min_{|v|=1} \int_{(y-x)\cdot v>0} \rho\, dy. \tag{93}$$

The Tukey depth is also called the half-space depth, and any point with the largest Tukey depth can be defined as the Tukey Depth Median. Tukey depth has been extended to graphs (Small, 1997) and metric spaces (Carrizosa, 1996), and at the populuation level has been recently connected to a nonstandard eikonal equation (Molina-Fructuoso and Murray, 2021). Other notions of data depth include the Monge-Kantorovich depth (Chernozhukov et al., 2017), depth for curves (de Micheaux et al., 2020), and data peeling (Calder and Smart, 2020; Calder et al., 2014, 2015; Calder, 2016, 2017; Bou-Rabee and Morfe, 2021; Cook and Calder, 2022). For a general survey on ordering of multivariate data, we refer to (Barnett, 1976; Liu et al., 1999).

We mention that when using density weighting in any of these approaches, the type of weighting that will be effective (i.e., directly proportional, or inversely) depends on the notion of median or data depth under consideration. For example, in the generalized geometric median, or barycenter problem (92) that we consider, the density weighting is performed in the computation of the quantity $d(x, y)$, which measures the distance between x and y. In this context, inverse density weighting, where paths in high density regions are very short and inexpensive, while paths in low density regions are longer, encourages the median to be placed in a high density region that can be quickly accessed from anywhere in the dataset, which is desirable in many applications. Conversely, if one were to re-weight the density in the Tukey depth (93), that is, replace ρ by $\rho^{-\alpha}$ for some exponent α, then it is important to ensure the weighting is proportional to the density (i.e., $\alpha < 0$).

5.2 Analysis of median and data depth

We first begin with a continuum analysis of the p-eikonal median and data depth. The continuum limit of the discrete p-eikonal median (26) is the geodesic geometric median

$$x_* \in \operatorname*{argmin}_{x\in\Omega} \int_\Omega d_{\rho^{-\alpha}}(y, \{x\})\rho(y)\, dy. \tag{94}$$

The associated depth is based on the distance to x_*, and is given by

$$\operatorname{depth}_\alpha(x) = \max_\Omega d_{\rho^{-\alpha}}(\cdot, x_*) - d_{\rho^{-\alpha}}(x, x_*). \tag{95}$$

We note that this notion of depth and median is only defined for distributions without compact support Ω. We expect that with some appropriate tail bounds, the case of unbounded distributions, like normal distributions, could be addressed as well.

We study here the case of a radial density $\rho(x) = \rho(|x|)$ that is radially decreasing on the unit ball $\Omega = B(0,1)$. In this case we expect the median to be the origin $x_* = 0$ for $\alpha > 0$. We are able to obtain a partial result for uniform densities.

Lemma 35. *If $\rho \equiv 1$ on $\Omega = B(0,1)$, then $x_* = 0$.*

Proof Since the ball is convex and $\rho \equiv 1$, we have that $d_{\rho^{-\alpha}}(x,y) = |x-y|$ for all $x, y \in B(0,1)$. Therefore

$$x_* \in \operatorname*{argmin}_{x\in B(0,1)} \int_{B(0,1)} |x-y|\, dy.$$

Let $x \neq 0$. We first note that

$$\int_{B(0,1)} |y|\, dy = \int_{B(x,1)} |x-y|\, dy = \int_{B(0,1)\cap B(x,1)} |x-y|\, dy + \int_{B(x,1)\setminus B(0,1)} |x-y|\, dy. \tag{96}$$

Since $x \neq 0$ and $|B(x,1) \setminus B(0,1)| = |B(0,1) \setminus B(x,1)|$ and $|x-y| < 1$ for $y \in B(x,1)$ we have

$$\int_{B(x,1)\setminus B(0,1)} |x-y|\, dy < \int_{B(x,1)\setminus B(0,1)} dy = \int_{B(0,1)\setminus B(x,1)} dy.$$

Since $1 < |x-y|$ for $y \in B(0,1) \setminus B(x,1)$ we obtain

$$\int_{B(x,1)\setminus B(0,1)} |x-y|\, dy < \int_{B(0,1)\setminus B(x,1)} |x-y|\, dy.$$

Substituting this into (96) yields

$$\int_{B(0,1)} |y|\, dy < \int_{B(0,1)\cap B(x,1)} |x-y|\, dy + \int_{B(0,1)\setminus B(x,1)} |x-y|\, dy = \int_{B(0,1)} |x-y|\, dy.$$

It follows that

$$0 = \operatorname*{argmin}_{x\in B(0,1)} \int_{B(0,1)} |x-y|\, dy,$$

which completes the proof. ■

Remark 36. *We expect that Lemma 35 holds for any radially decreasing density ρ on the unit ball $B(0,1)$ provided $\alpha \geq 0$, but it appears the proof would be substantially different than Lemma 35.*

If the median is at the origin, and ρ is radial, we can easily compute the depth function for any choice of density weighting.

Lemma 37. *Let $\alpha \in \mathbb{R}$. Assume $\rho(x) = \rho(|x|)$ is radial and $\Omega = B(0,1)$. If $x_* = 0$ then*

$$depth_\alpha(x) = \int_{1-|x|}^{1} \rho(r)^{-\alpha}\, dr.$$

Proof Since ρ is radial and decreasing, the shortest paths to the origin are straight lines. Indeed, let $x \in \Omega$, $x \neq 0$, and take any Lipschitz curve $\gamma : [0,1] \to \Omega$ with $\gamma(0) = 0$ and $\gamma(1) = x$. The density weighted length of this path is

$$\ell := \int_0^1 \rho(|\gamma(t)|)^{-\alpha} |\gamma'(t)|\, dt.$$

We now use the change of variables $r = |\gamma(t)|$, for which $|r'(t)| \leq |\gamma'(t)|$. The change of variables formula (Evans and Garzepy, 2018, Theorem 3.9) yields

$$\ell \geq \int_0^1 \rho(|\gamma(t)|)^{-\alpha} |r'(t)|\, dt = \int_0^{|x|} \left(\sum_{t\,:\,|\gamma(t)|=r} \rho(r)^{-\alpha} \right) dr \geq \int_0^{|x|} \rho(r)^{-\alpha}\, dr.$$

Since the integral on the right hand side is the density weighted length of the straight line $\gamma(t) = \frac{x}{|x|} t$, the claim is established.

It follows that

$$d_{\rho^{-\alpha}}(x, 0) = \int_0^{|x|} \rho(r)^{-\alpha}\, dr.$$

Hence $\max_\Omega d_{\rho^{-\alpha}}(\cdot, 0) = \int_0^1 \rho(r)^{-\alpha}\, dr$, which completes the proof. ∎

Remark 38. *Note in Lemma 37 that if we take $\rho \equiv 1$ then $depth_\alpha(x) = 1 - |x|$.*

Finally, we show that in one dimension the p-eikonal depth reduces to the usual notion of median, regardless of the choice of density weighting.

Lemma 39. *Assume that $d = 1$, $\Omega = (a, b)$, $\alpha \in \mathbb{R}$, and $\rho > 0$ on Ω. Then the p-eikonal median x_* is the median of ρ, that is, it holds that*

$$\int_a^{x_*} \rho(x)\, dx = \int_{x_*}^{b} \rho(x)\, dx.$$

Proof In this setting, for any $x, y \in [a, b]$ the density weighted distance is given by

$$d_{\rho^{-\alpha}}(y, \{x\}) = \int_{[x,y]} \rho(t)^{-\alpha}\, dt.$$

Therefore, the continuum p-ekkonal median problem (94) becomes

$$\min_{x\in[a,b]} g(x) := \int_a^b \int_{[x,y]} \rho(t)^{-\alpha}\, dt\, \rho(y)\, dy.$$

Differentiating in x we obtain

$$\begin{aligned} g'(x) &= \int_a^b \left(\mathbb{1}_{\{y\le x\}} \rho(x)^{-\alpha} - \mathbb{1}_{\{y>x\}} \rho(x)^{-\alpha} \right) \rho(y)\, dy \\ &= \rho(x)^{-\alpha} \left(\int_a^x \rho(y)\, dy - \int_x^b \rho(y)\, dy \right). \end{aligned}$$

Let x_m be the median of ρ, which satisfies

$$\int_a^{x_m} \rho(y)\, dy - \int_{x_m}^b \rho(y)\, dy = 0.$$

Then $g'(x_m) = 0$. For $x > x_m$ we clearly have $g'(x) > 0$, while for $x < x_m$ we have $g'(x) < 0$. This implies that x_m is the global minimizer of $g(x)$, and so $x_* = x_m$. ∎

5.3 Analysis of semi-supervised learning

In order to study the consistency of semi-supervised learning, we make a clusterability assumption on the density ρ. We assume there are k classes, represented by the open and connected sets $\Omega_1, \dots, \Omega_k \subset \Omega$, all of which are mutually disjoint. For each $j = 1, \dots, k$ we let

$$\rho_j = \min_{\overline{\Omega_j}} \rho,$$

and we set $\widetilde{\Omega} = \Omega \setminus \bigcup_{j=1}^k \Omega_j$ and

$$\delta = \max_{\widetilde{\Omega}} \rho.$$

We assume there are closed sets $\Gamma_j \subset \Omega_j$ for each $j = 1, \dots, k$ that correspond to the labeled data for each class. Then the continuum limit of the p-eikonal semi-supervised learning algorithm from Section 2.4.1 produces the predicted labels $\ell : \Omega \to \{1, \dots, k\}$ given by

$$\ell(x) = \operatorname*{argmin}_{1\le j\le k} d_{\rho^{-\alpha}}(x, \Gamma_j). \tag{97}$$

Definition 40. *We say that the classification is* asymptotically consistent *if for all $j = 1, \dots, k$ we have $\ell(x) = j$ for all $x \in \Omega_j$.*

Note that the definition of asymptotic consistency does not place any conditions on the label function in the space between classes $\widetilde{\Omega}$.

We define the Hausdorff distance

$$\mathcal{H}(\Gamma_j, \Omega_j) = \max_{x\in\overline{\Omega_j}} d_{\Omega_j}(x, \Gamma_j),$$

which measures how well the labeled set Γ_j covers the class Ω_j via geodesic distance on Ω_j. We note that for any $A \subset \Omega$, we take the definition of d_A to be $d_A(x,\Gamma) = d_f(x,\Gamma)$ where $f = \mathbb{1}_A$ is the indicator function of A, and d_f is defined in Section 3. Thus, the feasible paths for $d_A(x,\Gamma)$ can travel anywhere in Ω, but we only measure the length of the segments of the path that lie in A. We also define the separation of classes i and j by

$$\mathcal{S}(\Omega_i,\Omega_j) = \min\{d_{\widetilde{\Omega}}(x,y) \,:\, x \in \Omega_i \;\text{ and }\; y \in \Omega_j\}.$$

The separation $\mathcal{S}(\Omega_i,\Omega_j)$ is the length of the shortest path from a point in Ω_i to a point in Ω_j, where only the distance traveled in $\widetilde{\Omega}$ is counted.

We now define

$$\beta_{ij} = \frac{\delta^\alpha \mathcal{H}(\Gamma_j,\Omega_j)}{\rho_j^\alpha \mathcal{S}(\Omega_i,\Omega_j)}. \tag{98}$$

We will assume that $\Gamma_j \subsetneq \Omega_j$ for each j, so that $\beta_{ij} > 0$ for all $i \neq j$ (otherwise the classification of class j is trivial). As we shall see in the results below, our clusterability assumption relates to the smallness of β_{ij}. This includes measures of how well Γ_j covers Ω_j, the ratio of the background density δ to the class density ρ_j, and the separation between classes i and j.

Theorem 41. *Let $\alpha \geq 0$. If $\beta_{ij} < 1$ for all $i \neq j$, then the classification* (97) *is asymptotically consistent.*

Proof To show that the classification is asymptotically consistent, we need to show that for all $i \neq j$ we have

$$d_{\rho^{-\alpha}}(x,\Gamma_j) < d_{\rho^{-\alpha}}(x,\Gamma_i) \;\text{ for all } x \in \Omega_j. \tag{99}$$

Let $x \in \Omega_j$. Since $\rho \geq \rho_j$ on Ω_j we have

$$d_{\rho^{-\alpha}}(x,\Gamma_j) \leq \rho_j^{-\alpha} d_\Omega(x,\Omega_j) \leq \rho_j^{-\alpha}\mathcal{H}(\Gamma_j,\Omega_j).$$

Similarly, since $\rho \leq \delta$ in $\widetilde{\Omega}$ we have

$$d_{\rho^{-\alpha}}(x,\Gamma_i) \geq \delta^{-\alpha}\mathcal{S}(\Omega_i,\Omega_j).$$

Combining these two inequalities, we have that (99) holds provided

$$\delta^{-\alpha}\mathcal{S}(\Omega_i,\Omega_j) > \rho_j^{-\alpha}\mathcal{H}(\Gamma_j,\Omega_j)$$

for all $i \neq j$. Rearranging we obtain $\beta_{ij} < 1$, which completes the proof. ■

We now consider the inclusion of class priors. Given positive weights $s_1,\ldots,s_k$, the continuum limit of the class priors label decision (28) is given by

$$\ell(x) = \operatorname*{argmin}_{1\leq j\leq k}\{s_j d_{\rho^{-\alpha}}(x,\Gamma_j)\}. \tag{100}$$

Theorem 42. *Let $\alpha \geq 0$ and define*

$$[\beta]_* = \max_C \left(\prod_{(i,j)\in C} \beta_{ij} \right)^{\frac{1}{|C|}},$$

where the maximum is over all cycles of $\{1, 2, \ldots, k\}$. If $[\beta]_* < 1$, then there exists $s \in \mathbb{R}^k_+$ such that the classification (100) *is asymptotically consistent.*

Remark 43. *We note that $[\beta]_* \leq \max_{i\neq j} \beta_{ij}$, so Theorem 42 shows that the utilization of class priors leads to a weaker condition for asymptotic consistency. In the case of binary classification, $k = 2$, there is only one cycle $C = \{(1, 2), (2, 1)\}$ and we have*

$$[\beta]_* = \sqrt{\beta_{12}\beta_{21}}.$$

Thus, Theorem 42 shows that the class priors label decision (100) *with the optimal choice of s is asymptotically consistent for binary classification provided $\beta_{12}\beta_{21} < 1$, which allows, for example $\beta_{12} > 1$ and $\beta_{21} < 1$ (or vice versa). This is a much more relaxed condition compared to the consistency of the label decision* (97) *without class priors, which requires both $\beta_{12} < 1$ and $\beta_{21} < 1$. Thus, Theorem 42 shows how class priors are able to correct for poor separation between classes, poor choices of labeled training data, or low density clusters, provided there is another class with good clusterability properties to tradeoff with.*

The proof of Theorem 42 is based on an alternative characterization of $[\beta]_*$.

Proposition 44. *We have*

$$[\beta]_* = \min_{s\in\mathbb{R}^k_+} \max_{i\neq j} \{s_i^{-1} s_j \beta_{ij}\}. \tag{101}$$

Proof Let us define $F : \mathbb{R}^k_+ \to \mathbb{R}$ by

$$F(s) = \max_{i\neq j} \{s_i^{-1} s_j \beta_{ij}\}. \tag{102}$$

We first show that the minimum of F exists. Since only ratios of s appear, we may restrict to s with $s_1 = 1$. Set $\beta_{min} = \min_{i\neq j} \beta_{ij}$ and $\beta_{max} = \max_{i\neq j} \beta_{ij}$, and note that $\beta_{min} > 0$ by assumption. Then $F(s) \geq \beta_{min} s_j$ for all j. Since $\inf F \leq \beta_{max}$, we may also restrict to s such that $\beta_{min} s_j \leq \beta_{max}$, that is $s_j \leq \beta_{max}/\beta_{min}$. Likewise, we have $F(s) \geq \beta_{min} s_i^{-1}$ for all all i, so we may restrict to s with $\beta_{min} s_i^{-1} \leq \beta_{max}$, or $s_i \geq \beta_{min}/\beta_{max}$. Thus, we have reduced the problem to minimizing the continuous function F over a compact set, and so the minimum exists.

Let us write $F_* = \min_{s\in\mathbb{R}^k_+} F(s)$. Let $s \in \mathbb{R}^k_+$ be a minimizer of F. Let C be any cycle in the complete graph on $\{1, 2, \ldots, k\}$. Then since $s_i^{-1} s_j \beta_{ij} \leq F_*$ for all $i \neq j$ we have

$$\prod_{(i,j)\in C} s_i^{-1} s_j \beta_{ij} \leq F_*^{|C|}.$$

In the product on the left side, the weights s_i all cancel out, since C is a cycle, and so we have

$$F_* \geq \left(\prod_{(i,j)\in C} \beta_{ij} \right)^{\frac{1}{|C|}}. \tag{103}$$

Maximizing over C on the right hand side yields one direction of the proposition, that $F_* \geq [\beta]_*$.

To prove the other direction, for $s \in \mathbb{R}^k_+$ let us define

$$M(s) = \{(i,j) \,:\, i \neq j \;\text{ and }\; s_i^{-1} s_j \beta_{ij} = F_*\}.$$

For any minimizer s of F, we have $\#M(s) \geq 2$. Indeed, by the definition of F in (102), we must have at least one pair (i,j), $i \neq j$, with $s_i^{-1} s_j \beta_{ij} = F(s) = F_*$, since s is optimal, so $(i,j) \in M(s)$ and $\#M(s) \geq 1$. Assume now, by way of contradiction, that $\#M(s) = 1$. Then $(j,i) \notin M(s)$ and so $s_j^{-1} s_i \beta_{ji} < F_*$. Since $\#M(s) = 1$ we have $s_k^{-1} s_\ell \beta_{k\ell} < F_*$ for all other pairs $(k,\ell) \neq (i,j)$. Let $\varepsilon > 0$ and define $\widetilde{s} \in \mathbb{R}^k_+$ by $\widetilde{s}_i = s_i + \varepsilon$, $\widetilde{s}_j = s_j - \varepsilon$ and $\widetilde{s}_k = s_k$ for all $k \notin \{i,j\}$. For sufficiently small $\varepsilon > 0$ we have

$$\max_{i \neq j} \{ s_i^{-1} s_j \beta_{ij} \} < F_* = F(s),$$

which is a contradiction to the minimality of s. Therefore $\#M(s) \geq 2$.

We now select a minimizer s for which $M(s)$ contains the fewest number of edges (this minimizer need not be unique). We claim that $M(s)$ must contain a cycle. To see this, note that if $M(s)$ did not contain a cycle, then there would exist an edge $(i,j) \in M(s)$ such that $(j,k) \notin M(s)$ for all $k \neq j$. We can therefore decrease s_j slightly (similar to the argument above) to produce another minimizer $\widetilde{s}$ with $2 \leq \#M(\widetilde{s}) < \#M(s)$, which contradicts our selection of s. Therefore $M(s)$ must contain a cycle.

Let C be a cycle contained in $M(s)$. Then for each $(i,j) \in C$ we have $s_i^{-1} s_j \beta_{ij} = F_*$ and so

$$\prod_{(i,j)\in C} \beta_{ij} = \prod_{(i,j)\in C} s_i^{-1} s_j \beta_{ij} = F_*^{|C|},$$

which shows that $F_* \leq [\beta]_*$, and completes the proof. ∎

We now give the proof of Theorem 42.

Proof [Proof of Theorem 42] To show that the classification is asymptotically consistent, we need to show that there exist weights s_j such that for all $i \neq j$ we have

$$s_j d_{\rho^{-\alpha}}(x, \Gamma_j) < s_i d_{\rho^{-\alpha}}(x, \Gamma_i) \;\text{ for all } x \in \Omega_j. \tag{104}$$

Applying the same arguments as in the proof of Theorem 41, we find that (104) is equivalent to $s_i^{-1} s_j \beta_{ij} < 1$ for all $i \neq j$. If $[\beta]_* < 1$, then such weights exist, by Proposition 44, and the proof is complete. ∎

Remark 45. *We briefly mention that all of the results in this section can be extended to hold for the finite sample size p-eikonal learning problem, with high probability and additional error terms, due to the quantitative convergence results given in Theorems 22 and 23. To avoid additional technicalities, we leave such considerations to future work.*

Remark 46. *We also mention that all of our results in this section are in the setting of a dataset that is highly clusterable (due to the assumption that $\Omega_i \cap \Omega_j = \varnothing$), so that only a finite number of labels are required to obtain asymptotic classification consistency in the continuum limit. In the setting where there is overlap between classes, and the dataset is not very well-clusterable, then we expect that the number of labels required for asymptotic consistency of classification should grow to infinity as $n \to \infty$, at some appropriate rate. We leave investigations of these higher label rate problems to future work.*

6. Numerical experiments

We present here some numerical experiments with real datasets. All code for the experiments is available online[6] and uses the GraphLearning Python package (Calder, 2022). In all experiments we solved the graph p-eikonal equation (20) with the fast marching solver described in Section 2.3.2, implemented in the C programming language. The rest of this section is broken up into data depth experiments in Section 6.1 and semi-supervised learning experiments in Section 6.2.

6.1 Data depth

We consider the MNIST dataset of handwritten digits (LeCun et al., 1998) and the Fashion-MNIST dataset (Xiao et al., 2017), which is a drop-in replacement for MNIST consisting of 10 classes of clothing items. Each dataset has 70,000 grayscale images of size 28×28 pixels. For both datasets we restricted the computations of data depth to each individual class, which consists of about 7000 datapoints per class. We constructed the graph by connecting each image to its K-nearest neighbors with Gaussian weights given by

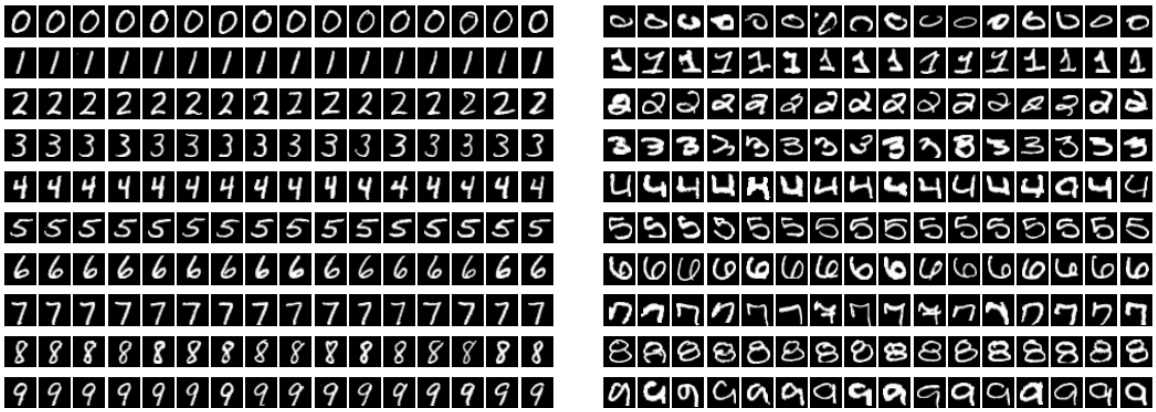
$$w_{ij} = \exp\left(-\frac{4|x_i - x_j|^2}{d_K(x_i)^2}\right), \tag{105}$$

where x_i represents the pixel values for image i, and $d_K(x_i)$ is the distance between x_i and its K^{th} nearest neighbor. We used $K = 20$ in all experiments. The weight matrix was then symmetrized by replacing W with $W + W^T$.

We computed the p-eikonal median via the definition (26) with $p = 1$ and $\alpha = 2$. For the density estimator $\hat{\rho}$ we used a k-nearest neighbor density estimator with $k = 30$. To speed up the computation of (26), we computed the minimum in (26) over 5% of the nodes in each class, chosen at random. This takes about 5 minutes to compute for each dataset (30 seconds per class), which includes the time for the k-nearest neighbor search.

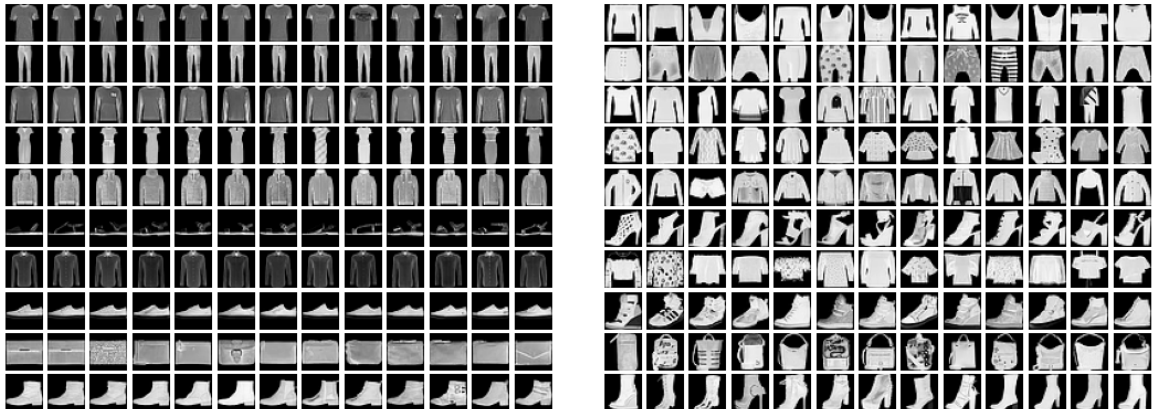
In Figures 7 and 8 we show the deepest images (i.e., the medians) and the shallowest images (i.e., outliers) from each class for the MNIST and FashionMNIST datasets. We can see that the deepest handwritten digits are very clean and self-consistent, while the

6. `https://github.com/jwcalder/peikonal`



(a) Deepest images (median) (b) Shallowest images (outliers)

Figure 7: Comparison of deepest (median) images to shallowest (outlier) images from each MNIST digit.



(a) Deepest images (median) (b) Shallowest images (outliers)

Figure 8: Comparison of deepest (median) images to shallowest (outlier) images from each FashionMNIST class.

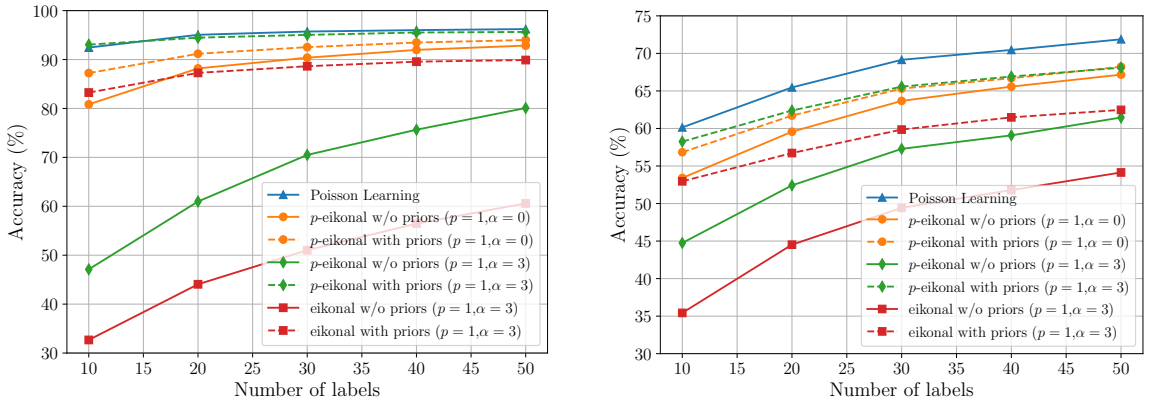
shallowest do appear visually to be outliers. For FashionMNIST the deepest images are again self-similar and very plain, while the shallowest images tend to be more varied and have patterns on the clothing items. Finally, in Figure 9 we show paths through each class from the shallowest point to the deepest point, following the gradient descent path construction from Section 2.3.3.

6.2 Semi-supervised learning

We tested the p-eikonal equation for semi-supervised learning at very low label rates with $p = 1$. In addition to MNIST and FashionMNIST, we also tested on CIFAR-10 (Krizhevsky

(a) MNIST (b) FashionMNIST

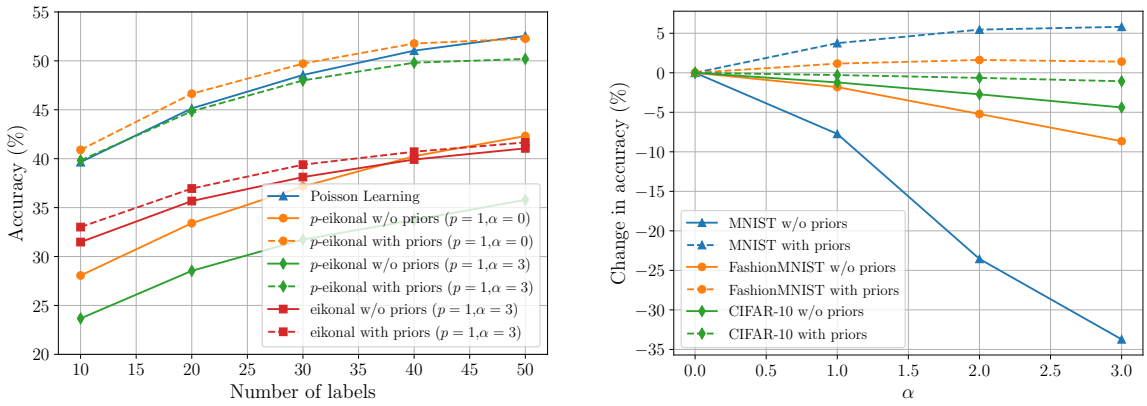
Figure 9: Paths from shallowest point to median for each class computed with the gradient descent method from Section 2.3.3.



(a) MNIST (b) FashionMNIST

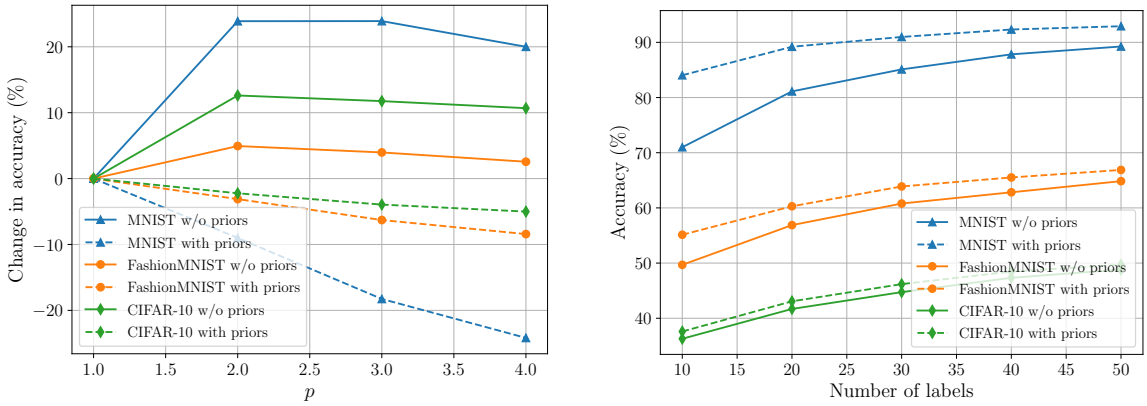
Figure 10: Comparison of the p-eikonal equation with $p = 1$ for semi-supervised image classification to Poisson learning (Calder et al., 2020a) and the eikonal equation (17).

et al., 2009). To build good quality graphs for classification, we cannot use the pixel-wise differences that we did for data depth in Section 6.1. Instead we follow the methods from (Calder et al., 2020a) and trained autoencoders to extract important features from the data. For MNIST and FashionMNIST, we used variational autoencoders, similar to (Kingma and Welling, 2014), while for CIFAR-10 we used the AutoEncoding Transformations architecture from (Zhang et al., 2019). After training the autoencoders we built K-nearest neighbor graphs with weights given by (105) over the latent variables using the angular similarity with $K = 20$ neighbors. We again used a k-nearest neighbor density estimator with $k = 30$ neighbors for the reweighting. We refer to (Calder et al., 2020a) for more details about the



(a) CIFAR-10 (b) Accuracy vs α

Figure 11: (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 α is adjusted.



(a) Accuracy vs p (b) $p = 2, \alpha = 3$

Figure 12: Comparison of how the classification accuracy depends on the exponent p in the p-eikonal equation. In both experiments we used density weighting with $\alpha = 3$.

autoencoder graph construction, which was also used successfully in another recent work (Miller et al., 2022). After the graphs have been constructed, the classification results on any of the 3 datasets, which requires solving 10 p-eikonal equations, takes a few seconds to run the classification for each trial.

We ran 100 trials at 1 label per class up to 5 labels per class, randomly choosing different labeled data for each trial. We compared against Poisson learning (Calder et al., 2020a) and the graph distance eikonal equation (17) with the same density reweighting schemes. We tested the p-eikonal and eikonal equations with and without class priors, as described in Section 2.4.2. Figure 10 shows the results for MNIST and FashionMNIST, while Figure

11a shows the results on CIFAR-10. We see that with class priors, p-eikonal learning is comparable to Poisson learning on MNIST, slightly worse on FashionMNIST and slightly better on CIFAR-10. We also see that p-eikonal offers a significant improvement over the shortest path based eikonal classifier, even though we applied the same density reweighting to both. Since the asymptotic consistency results from Section 5.3 would hold equally well for the density reweighted eikonal equation, we attribute the improved results to the robustness properties of the p-eikonal equation (see Theorem 14) to perturbations in graphs, which are common in real data.

In Figure 11b we show how the accuracy changes for each dataset as the density exponent is increased. We find a quite surprising result here; without class priors the accuracy actually *decreases* when density reweighting is used. A possible explanation of this is that the density reweighting makes the optimal classification thresholds more dependent on the cluster geometry and density. Indeed, it is only with the addition of class priors that the density reweighting can increase the accuracy of the classifier. This is true across all datasets and validates our theoretical findings in Theorem 42 that class priors can effectively make use of density reweighting to improve classification results.

Finally, in Figure 12a we show how the accuracy changes as the exponent p in the p-eikonal equation is changed (here, $\alpha = 3$). All our previous experiments were with $p = 1$, and we find another surprising result here; the classification accuracy improves up to $p = 2$ without class priors, but is monotonically decreasing when utilizing class priors. This may simply be due to the fact that the classification accuracy is already very high with class priors, and very low without. Indeed, in Figure 12b we show the accuracy for $p = 2$ and $\alpha = 3$ for each dataset, and both with and without class priors. We see that even though $p = 2$ is better for the classifiers without class priors, the incorporation of class priors still improves the accuracy significantly.

7. Conclusion

We introduced and studied a family of graph-based distance-type equations called the p-eikonal equation. We showed that the p-eikonal equation for $p = 1$ is a robust estimator of the geodesic density weighted path distance on the underlying Euclidean space, compared to the standard shortest-path graph distance. We proved that, while the p-eikonal equation is not a distance function on a graph, it has similar properties and its continuum limit recovers the geodesic density weighted distance on the underlying Euclidean space, with quantitative convergence rates. We used the continuum limit theory to prove asymptotic consistency of data depth and semi-supervised learning with the p-eikonal equation and then gave some experiments with real data on the MNIST, FashionMNIST, and CIFAR-10 datasets.

Acknowledgments and Disclosure of Funding

The authors thank the Institute for Mathematics and its Applications (IMA), where part of this work was conducted. JC acknowledges funding from NSF grant DMS:1944925, the Alfred P. Sloan foundation, and a McKnight Presidential Fellowship.

A. Concentration of measure

We recall here some useful concentration of measure results, the proofs of which can be found in (Boucheron et al., 2013).

Theorem 47. *(Bernstein inequality) Let $x_1, x_2, \dots, x_n$ be a sequence of i.i.d real-valued random variables with finite expectation $\mu = \mathbb{E}(x_i)$ and variance $\sigma^2 = Var(x_i)$, and write $S_n := \frac{1}{n}\sum_{i=1}^n x_i$. Assume there exists $b > 0$ such that $|x_i - \mu| \leq b$ almost surely. Then for any $t > 0$ we have*

$$\mathbb{P}(S_n - \mu \geq t) \leq \exp\Big(- \frac{nt^2}{2(\sigma^2 + \frac{bt}{3})}\Big) \tag{106}$$

Theorem 48. *(Chernoff bounds) Let $x_1, x_2, \dots, x_n$ be a sequence of i.i.d Bernoulli random variables with parameter $p \in [0,1]$. Then for any $\delta > 0$ we have*

$$\mathbb{P}\Big(\sum_{i=1}^n x_i \geq (1+\delta)np\Big) \leq \exp\Big(- \frac{np\delta^2}{2(1+\frac{1}{3}\delta)}\Big) \tag{107}$$

and for any $0 \leq \delta < 1$ we have

$$\mathbb{P}\Big(\sum_{i=1}^n x_i \leq (1-\delta)np\Big) \leq \exp\Big(- \frac{1}{2}np\delta^2\Big) \tag{108}$$

B. Technical proofs

We include here some technical, but elementary, proofs from the paper.

Proof [Proof of Theorem 7] Since H admits comparison, there is at most one solution of (4), so we only have to establish existence. We use the Perron method. Let $\mathcal{F}$ be the set of all $v \in F(X)$ such that

$$H(\nabla_X v(x_i), v(x_i), x_i) \leq 0 \quad \text{for all } x_i \in X \setminus \Gamma$$

and $v = g$ on Γ. The set $\mathcal{F}$ is nonempty, since $\varphi \in \mathcal{F}$. Define the Perron function

$$u(x_i) = \sup\{v(x_i) \,:\, v \in \mathcal{F}\}.$$

Since H admits comparison, we have $v \leq \psi$ for all $v \in \mathcal{F}$, and so $\varphi \leq u \leq \psi$.

We now claim that

$$H(\nabla_X u(x_i), u(x_i), x_i) \leq 0 \quad \text{for all } x_i \in X \setminus \Gamma.$$

To see this, let $x_i \in X \setminus \Gamma$ and let $\varepsilon > 0$. There exist $v \in \mathcal{F}$ such that $u(x_i) \leq v(x_i) + \varepsilon$. By definition we have $u(x_j) \geq v(x_j)$ for all j, and so $\nabla_X u(x_i) - \varepsilon \mathbb{1} \leq \nabla_X v(x_i)$. Therefore

$$0 \geq H(\nabla_X v(x_i), v(x_i), x_i) \geq H(\nabla_X u(x_i) - \varepsilon\mathbb{1}, u(x_i) - \varepsilon, x_i).$$

Sending $\varepsilon \to 0$ and using continuity of H establishes the claim.

We now claim that

$$H(\nabla_X u(x_i), u(x_i), x_i) \geq 0 \quad \text{for all } x_i \in X \setminus \Gamma,$$

which will complete the proof. Assume, by way of contradiction, that

$$H(\nabla_X u(x_i), u(x_i), x_i) < 0 \quad \text{for some } x_i \in X \setminus \Gamma.$$

Let $\varepsilon > 0$ and define $u_\varepsilon \in F(X)$ by $u_\varepsilon(x_j) = u(x_j)$ for $j \neq i$, and $u_\varepsilon(x_i) = u(x_i) + \varepsilon$. By continuity of H, there is a sufficiently small $\varepsilon > 0$ so that

$$H(\nabla_X u_\varepsilon(x_i), u_\varepsilon(x_i), x_i) \leq 0.$$

Furthermore, for any $j \neq i$, we have $u_\varepsilon(x_j) = u(x_j)$ and $\nabla_X u_\varepsilon(x_j) \leq \nabla_X u(x_j)$. Since H is monotone we find that

$$H(\nabla_X u_\varepsilon(x_j), u_\varepsilon(x_j), x_j) \leq H(\nabla_X u(x_j), u(x_j), x_j) \leq 0$$

for $j \neq i$ with $x_j \in X \setminus \Gamma$. Therefore $u_\varepsilon \in \mathcal{F}$, which is a contradiction (since $u_\varepsilon(x_i) > u(x_i)$), establishing the claim and completing the proof. ■

Proof [Proof of Theorem 19] The proof uses the following dynamic programming principle

$$u(x) = \min_{y \in \partial B(x,r) \cap \overline{\Omega}} \{u(y) + d_f(x,y)\},$$

which holds provided $B(x,r) \subset \Omega \setminus \Gamma$ and is immediate to verify. Rearranging the dynamic programming principle we obtain

$$\max_{y \in \partial B(x,r) \cap \overline{\Omega}} \{u(x) - u(y) - d_f(x,y)\} = 0. \tag{109}$$

Since f is Lipschitz continuous, we have

$$d_f(x,y) = f(x)|x-y| + \mathcal{O}(|x-y|^2),$$

which, when substituted above, yields

$$\max_{y \in \partial B(x,r) \cap \overline{\Omega}} \left\{ \frac{u(x) - u(y)}{r} \right\} = f(x) + \mathcal{O}(r). \tag{110}$$

We now prove the subsolution property. Let $x \in \Omega \setminus \Gamma$ and let $\varphi \in C^\infty(\mathbb{R}^d)$ such that $u - \varphi$ has a local maximum at x. Then for $r > 0$ sufficiently small we have $B(x,r) \subset \Omega \setminus \Gamma$ and

$$u(x) - \varphi(x) \geq u(y) - \varphi(y) \quad \text{for all } y \in B(x,r).$$

Rearranging we have

$$u(x) - u(y) \geq \varphi(x) - \varphi(y) \quad \text{for all } y \in B(x,r).$$

Plugging this into (110) yields

$$\max_{y\in\partial B(x,r)}\left\{\frac{\varphi(x)-\varphi(y)}{r}\right\}\leq f(x)+\mathcal{O}(r).$$

Notice the maximum is over only $\partial B(x,r)$, since $B(x,r)\subset\Omega\setminus\Gamma$. Sending $r\to 0$ yields $|\nabla\varphi(x)|\leq f(x)$, which is exactly the subsolution property.

To prove the supersolution property, let $x\in\overline{\Omega}\setminus\Gamma$ and let $\varphi\in C^\infty(\mathbb{R}^d)$ such that $u-\varphi$ has a local minimum at x. As above, this means that

$$u(x)-u(y)\leq\varphi(x)-\varphi(y)\ \text{ for all } y\in B(x,r)\cap\overline{\Omega}.$$

For $r>0$ small enough $B(x,r)\subset\overline{\Omega}\setminus\Gamma$, and so we can substitute this into (110) to obtain

$$\max_{y\in\partial B(x,r)\cap\overline{\Omega}}\left\{\frac{\varphi(x)-\varphi(y)}{r}\right\}\geq f(x)+\mathcal{O}(r).$$

By enlarging the domain in the maximum above, we obtain

$$\max_{y\in\partial B(x,r)}\left\{\frac{\varphi(x)-\varphi(y)}{r}\right\}\geq f(x)+\mathcal{O}(r).$$

We now send $r\to 0$ to obtain $|\nabla\varphi(x)|\geq f(x)$, which completes the proof. ■

Proof [Proof of Proposition 31] We note that the inequality (35) can be restated as

$$B(x,r)\subset B_\Omega(x,r+Cr^2)\ \text{ and }\ B_\Omega(x,r)\subset B(x,r). \tag{111}$$

For $r>0$ sufficiently small, so that $Cr\leq\frac{1}{2}$, the first inclusion above implies that

$$B_\Omega(x,r)\supset B(x,r-Cr^2)\supset B(x,\tfrac{r}{2}). \tag{112}$$

Since the boundary $\partial\Omega$ is $C^{1,1}$, there exists $v\in\mathbb{R}^d$ with $|v|=1$ and $c>0$ such that

$$B(x,\tfrac{r}{2})\cap\Omega\supset\{y\in B(x,\tfrac{r}{2})\,:\,(y-x)\cdot v\geq cr^2\}.$$

For r smaller, so that $cr\leq\frac{1}{4}$ as well, we have

$$\begin{aligned}|B_\Omega(x,r)\cap\Omega|&\geq|B(x,\tfrac{r}{2})\cap\Omega|\\&\geq|\{y\in B(x,\tfrac{r}{2})\,:\,(y-x)\cdot v\geq\tfrac{r}{4}\}|\\&=\left(\frac{r}{2}\right)^d|\{z\in B(0,1)\,:\,z\cdot v\geq\tfrac{1}{2}\}|\\&=c_dr^d\end{aligned}$$

where

$$c_d:=\frac{1}{2^d}\int_{B(0,1)\cap\{z_1\geq\frac{1}{2}\}}dx.$$

We finally compute

$$\begin{aligned} c_d &= \frac{1}{2^d}\int_{\frac{1}{2}}^1 \omega_{d-1}(1-z_1^2)^{\frac{d-1}{2}}\,dz \\ &\geq \frac{\omega_{d-1}}{2^d}\int_{\frac{1}{2}}^1 z_1(1-z_1^2)^{\frac{d-1}{2}}\,dz \\ &= -\frac{\omega_{d-1}}{2^d(d+1)}(1-z_1^2)^{\frac{d+1}{2}}\Big|_{\frac{1}{2}}^1 \\ &= \frac{\omega_{d-1}}{2^d(d+1)}\left(\frac{3}{4}\right)^{\frac{d+1}{2}}. \end{aligned}$$

Applying the lower bound $\frac{3}{4} \geq \frac{1}{2}$ above to simplify the constant completes the proof. ■

References

Morteza Alamgir and Ulrike Von Luxburg. Shortest path distance in random k-nearest neighbor graphs. *arXiv preprint arXiv:1206.6381*, 2012.

Martino Bardi, Italo Capuzzo Dolcetta, et al. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*, volume 12. Springer, 1997.

Vic Barnett. The ordering of multivariate data. *Journal of the Royal Statistical Society: Series A (General)*, 139(3):318–344, 1976.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

Avleen S Bijral, Nathan Ratliff, and Nathan Srebro. Semi-supervised learning with density based distances. *arXiv preprint arXiv:1202.3702*, 2012.

Karsten M Borgwardt and Hans-Peter Kriegel. Shortest-path kernels on graphs. In *Fifth IEEE international conference on data mining (ICDM'05)*, pages 8–pp. IEEE, 2005.

Ahmed Bou-Rabee and Peter S Morfe. Hamilton-Jacobi scaling limits of pareto peeling in 2d. *arXiv preprint arXiv:2110.06016*, 2021.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Leon Bungert, Jeff Calder, and Tim Roith. Uniform convergence rates for Lipschitz learning on graphs. *IMA Journal of Numerical Analysis*, 2022.

Jeff Calder. A direct verification argument for the Hamilton-Jacobi equation continuum limit of nondominated sorting. *Nonlinear Analysis Series A: Theory, Methods, & Applications*, 141:88–108, 2016.

Jeff Calder. Numerical schemes and rates of convergence for the Hamilton-Jacobi equation continuum limit of nondominated sorting. *Numerische Mathematik*, 137(4):819–856, 2017.

Jeff Calder. Lecture notes on viscosity solutions. *Online Lecture Notes*, 2018a.

Jeff Calder. The game theoretic p-Laplacian and semi-supervised learning with few labels. *Nonlinearity*, 32(1), 2018b.

Jeff Calder. Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data. *SIAM Journal on Mathematics of Data Science*, 1:780–812, 2019.

Jeff Calder. GraphLearning Python Package. `doi:10.5281/zenodo.5850940`, 2022. `https://github.com/jwcalder/GraphLearning`.

Jeff Calder and N. García Trillos. Improved spectral convergence rates for graph Laplacians on ε-graphs and k-NN graphs. *Applied and Computational Harmonic Analysis*, 60:123–175, 2022.

Jeff Calder and Dejan Slepčev. Properly-weighted graph Laplacian for semi-supervised learning. *Applied Mathematics and Optimization: Special Issue on Optimization in Data Science*, 82:1111–1159, 2019.

Jeff Calder and Charles K. Smart. The limit shape of convex hull peeling. *Duke Mathematical Journal*, 169(11):2079–2124, 2020.

Jeff Calder, Selim Esedoḡlu, and Alfred O Hero. A Hamilton-Jacobi equation for the continuum limit of non-dominated sorting. *SIAM Journal on Mathematical Analysis*, 46(1): 603–638, 2014.

Jeff Calder, Selim Esedoḡlu, and Alfred O Hero. A PDE-based approach to nondominated sorting. *SIAM Journal on Numerical Analysis*, 53(1):82–104, 2015.

Jeff Calder, Brendan Cook, Matthew Thorpe, and Dejan Slepčev. Poisson Learning: Graph based semi-supervised learning at very low label rates. *Proceedings of the 37th International Conference on Machine Learning, PMLR*, 119:1306–1316, 2020a.

Jeff Calder, Dejan Slepčev, and Matthew Thorpe. Rates of convergence for Laplacian semi-supervised learning with low labeling rates. *arXiv preprint arXiv:2006.02765*, 2020b.

Jeff Calder, N. García Trillos, and Marta Lewicka. Lipschitz regularity of graph Laplacians on random data clouds. *SIAM Journal on Mathematical Analysis*, 54(1):1169–1222, 2022a.

Jeff Calder, Sangmin Park, and Dejan Slepčev. Boundary estimation from point clouds: Algorithms, guarantees and applications. *Journal of Scientific Computing*, 92(2):1–59, 2022b.

Italo Capuzzo-Dolcetta and P-L Lions. Hamilton-Jacobi equations with state constraints. *Transactions of the American Mathematical Society*, 318(2):643–683, 1990.

Emilio Carrizosa. A characterization of halfspace depth. *Journal of multivariate analysis*, 58(1):21–26, 1996.

Olivier Chapelle and Alexander Zien. Semi-supervised classification by low density separation. In *International workshop on artificial intelligence and statistics*, pages 57–64. PMLR, 2005.

Victor Chernozhukov, Alfred Galichon, Marc Hallin, and Marc Henry. Monge–kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, 45(1):223–256, 2017.

Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

Brendan Cook and Jeff Calder. Rates of convergence for the continuum limit of nondominated sorting. *SIAM Journal on Mathematical Analysis*, 54(1):872–911, 2022.

Michael G Crandall and Pierre-Louis Lions. Viscosity solutions of hamilton-jacobi equations. *Transactions of the American mathematical society*, 277(1):1–42, 1983.

Michael G Crandall, Lawrence C Evans, and P-L Lions. Some properties of viscosity solutions of hamilton-jacobi equations. *Transactions of the American Mathematical Society*, 282(2):487–502, 1984.

Pierre Lafaye de Micheaux, Pavlo Mozharovskyi, and Myriam Vimond. Depth for curve data and applications. *Journal of the American Statistical Association*, pages 1–17, 2020.

Xavier Desquesnes and Abderrahim Elmoataz. Nonmonotonic front propagation on weighted graphs with applications in image processing and high-dimensional data classification. *IEEE Journal of Selected Topics in Signal Processing*, 11(6):897–907, 2017.

Xavier Desquesnes, Abderrahim Elmoataz, and Olivier Lézoray. Eikonal equation adaptation on weighted graphs: fast geometric diffusion process for local and non-local image and data processing. *Journal of Mathematical Imaging and Vision*, 46(2):238–257, 2013.

Ahmed El Alaoui, Xiang Cheng, Aaditya Ramdas, Martin J Wainwright, and Michael I Jordan. Asymptotic behavior of ℓ_p-based Laplacian regularization in semi-supervised learning. In *Conference on Learning Theory*, pages 879–906, 2016.

Lawrence C Evans. *Partial differential equations*. American mathematical society, 2 edition, 2010.

Lawrence C Evans and Ronald F Garzepy. *Measure theory and fine properties of functions*. Routledge, 2018.

P Thomas Fletcher, Suresh Venkatasubramanian, and Sarang Joshi. The geometric median on riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1): S143–S152, 2009.

Mauricio Flores, Jeff Calder, and Gilad Lerman. Analysis and algorithms for Lp-based semi-supervised learning on graphs. *Applied and Computational Harmonic Analysis*, 60: 77–122, 2022.

Nicolás García Trillos, Moritz Gerlach, Matthias Hein, and Dejan Slepčev. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator. *Foundations of Computational Mathematics*, 20(4):827–887, 2020.

Franca Hoffmann, Bamdad Hosseini, Assad A Oberai, and Andrew M Stuart. Spectral analysis of weighted laplacians arising in data clustering. *Applied and Computational Harmonic Analysis*, 56:189–249, 2022.

Sung Jin Hwang, Steven B Damelin, and Alfred O Hero III. Shortest path through random points. *The Annals of Applied Probability*, 26(5):2791–2823, 2016.

Matt Jacobs, Ekaterina Merkurjev, and Selim Esedolu. Auction dynamics: A volume constrained MBO scheme. *Journal of Computational Physics*, 354:288–310, 2018.

Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. In *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 2014.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. *Citeseer*, 2009.

Rasmus Kyng, Anup Rao, Sushant Sachdeva, and Daniel A Spielman. Algorithms for Lipschitz learning on graphs. In *Conference on Learning Theory*, pages 1190–1223, 2015.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Marta Lewicka and Yuval Peres. Which domains have two-sided supporting unit spheres at every boundary point? *Expositiones Mathematicae*, 38(4):548–558, 2020.

Anna Little, Daniel McKenzie, and James M Murphy. Balancing geometry and density: Path distances on high-dimensional data. *SIAM Journal on Mathematics of Data Science*, 4(1):72–99, 2022.

Anna V Little, Mauro Maggioni, and James M Murphy. Path-based spectral clustering: Guarantees, robustness to outliers, and fast algorithms. *Journal of machine learning research*, 21, 2020.

Regina Y Liu, Jesse M Parelius, and Kesar Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh). *The annals of statistics*, 27(3):783–858, 1999.

Xiaoyi Mai and Romain Couillet. Random matrix-inspired improved semi-supervised learning on graphs. In *International Conference on Machine Learning*, 2018a.

Xiaoyi Mai and Romain Couillet. A random matrix analysis and improvement of semi-supervised learning for large dimensional data. *The Journal of Machine Learning Research*, 19(1):3074–3100, 2018b.

Juan J Manfredi, Adam M Oberman, and Alexander P Sviridov. Nonlinear elliptic partial differential equations and p-harmonic functions on graphs. *Differential Integral Equations*, 28(1-2):79–102, 2015.

Kevin Miller, Xoaquin Baca, Jack Mauro, Jason Setiadi, Zhan Shi, Jeff Calder, and Andrea Bertozzi. Graph-based active learning for semi-supervised classification of SAR data. *SPIE Defense and Commercial Sensing: Algorithms for Synthetic Aperture Radar Imagery XXIX*, 12095, 2022.

Stanislav Minsker. Geometric median and robust estimation in banach spaces. *Bernoulli*, 21(4):2308–2335, 2015.

Martin Molina-Fructuoso and Ryan Murray. Tukey depths and Hamilton-Jacobi differential equations. *arXiv:2104.01648*, 2021.

Martin Molina-Fructuoso and Ryan Murray. Eikonal depth: an optimal control approach to statistical depths. *arXiv:2201.05274*, 2022.

Amit Moscovich, Ariel Jaffe, and Boaz Nadler. Fast semi-supervised regression: a geodesic nearest neighbor approach. *Online*, 2016. `https://mosco.github.io/geodesicknn/geodesic_knn.pdf`.

Boaz Nadler, Nathan Srebro, and Xueyuan Zhou. Semi-supervised learning with the graph laplacian: The limit of infinite unlabelled data. *Advances in neural information processing systems*, 22:1330–1338, 2009.

Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.

Adam M Oberman. Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton–jacobi equations and free boundary problems. *SIAM Journal on Numerical Analysis*, 44(2):879–895, 2006.

Mathew Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003.

Alessandro Rozza, Mario Manzo, and Alfredo Petrosino. A novel graph-based fisher kernel method for semi-supervised learning. In *2014 22nd International Conference on Pattern Recognition*, pages 3786–3791. IEEE, 2014.

James A Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1595, 1996.

Zuoqiang Shi, Stanley Osher, and Wei Zhu. Weighted nonlocal Laplacian on interpolation from sparse data. *Journal of Scientific Computing*, 73(2-3):1164–1177, 2017.

Dejan Slepčev and Matthew Thorpe. Analysis of p-Laplacian regularization in semisupervised learning. *SIAM Journal on Mathematical Analysis*, 51(3):2085–2120, 2019.

Christopher G Small. Multidimensional medians arising from geodesics on graphs. *The Annals of Statistics*, pages 478–494, 1997.

Vinh-Thong Ta, Abderrahim Elmoataz, and Olivier Lézoray. Adaptation of eikonal equation over weighted graph. In *International conference on scale space and variational methods in computer vision*, pages 187–199. Springer, 2009.

Vinh-Thong Ta, Abderrahim Elmoataz, and Olivier Lézoray. Nonlocal PDEs-based morphology on weighted graphs for image and data processing. *IEEE transactions on Image Processing*, 20(6):1504–1516, 2010.

Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

John W Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, volume 2, pages 523–531, 1975.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Yiding Yang, Xinchao Wang, Mingli Song, Junsong Yuan, and Dacheng Tao. Spagan: Shortest path graph attention network. *arXiv preprint arXiv:2101.03464*, 2021.

Amber Yuan, Jeff Calder, and Braxton Osting. A continuum limit for the PageRank algorithm. *European Journal of Applied Mathematics*, 2021.

Liheng Zhang, Guo-Jun Qi, Liqiang Wang, and Jiebo Luo. Aet vs. aed: Unsupervised representation learning by auto-encoding transformations rather than data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2547–2555, 2019.

Xueyuan Zhou and Mikhail Belkin. Semi-supervised learning by higher order regularization. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 892–900, 2011.

Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the 20th International Conference on Machine learning (ICML-03)*, pages 912–919, 2003.