Applied Mathematics & Optimization
https://doi.org/10.1007/s00245-019-09637-3

Properly-Weighted Graph Laplacian for Semi-supervised Learning

Jeff Calder[1] · **Dejan Slepčev**[2]

Abstract
The performance of traditional graph Laplacian methods for semi-supervised learning degrades substantially as the ratio of labeled to unlabeled data decreases, due to a degeneracy in the graph Laplacian. Several approaches have been proposed recently to address this, however we show that some of them remain ill-posed in the large-data limit. In this paper, we show a way to correctly set the weights in Laplacian regularization so that the estimator remains well posed and stable in the large-sample limit. We prove that our semi-supervised learning algorithm converges, in the infinite sample size limit, to the smooth solution of a continuum variational problem that attains the labeled values continuously. Our method is fast and easy to implement.

Keywords Semi-supervised learning · Label propagation · Asymptotic consistency · PDEs on graphs · Gamma-convergence

Mathematics Subject Classification 49J55 · 35J20 · 35B65 · 62G20 · 65N12

1 Introduction

For many applications of machine learning, such as medical image classification and speech recognition, labeling data requires human input and is expensive [13], while unlabeled data is relatively cheap. *Semi-supervised* learning aims to exploit this dichotomy by utilizing the geometric or topological properties of the unlabeled data, in conjunction with the labeled data, to obtain better learning algorithms. A significant portion of the semi-supervised literature is on *transductive* learning, whereby a func-

✉ Jeff Calder
jcalder@umn.edu

Dejan Slepčev
slepcev@math.cmu.edu

[1] Department of Mathematics, University of Minnesota, Minneapolis, USA

[2] Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, USA

Published online: 07 December 2019

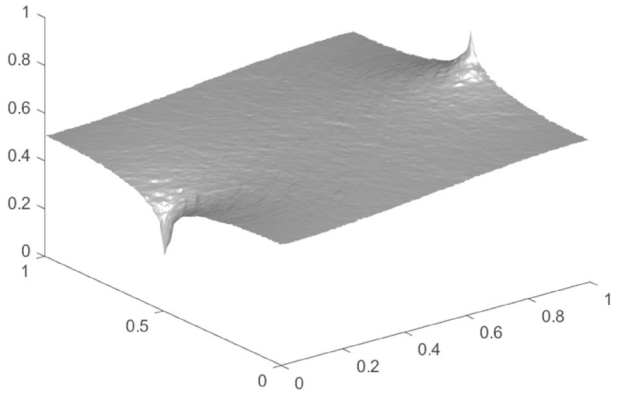


Fig. 1 Example of the degeneracy of graph Laplacian learning with few labels. The graph is a sequence of $n = 10^5$ *i.i.d.* random variables drawn from the unit box $[0, 1]^2$ in $\mathbb{R}^2$, and two labels are given $g(0, 0.5) = 0$ and $g(1, 0.5) = 1$

tion is learned only at the unlabeled points, and not as a parameterized function on an ambient space. In the transductive setting, graph based algorithms, such the graph Laplacian-based learning pioneered by [55], are widely used and have achieved great success [3,26,27,46–48,51–54].

Using graph Laplacians to propagate information from labeled to unlabeled points is one of the earliest and most popular approaches [55]. The constrained version of the graph Laplacian learning problem is to minimize over all $u : \mathcal{X} \to \mathbb{R}$

$$\begin{aligned} GL(u) &= \sum_{x,y \in \mathcal{X}} w_{xy}(u(x) - u(y))^2 \\ \text{subject to constraint} \quad & u(x) = g(x) \text{ for all } x \in \Gamma \end{aligned} \tag{1}$$

where the data points $\mathcal{X}$ form the vertices of a graph with edge weights w_{xy} and $\Gamma \subset \mathcal{X}$ are the labeled nodes with label function $g : \Gamma \to \mathbb{R}$. The minimizer u of (1) is the learned function, which extends the given labels g on Γ to the remainder of the graph. In classification contexts, the values of u are often rounded to the nearest label. The method amounts to minimizing a Dirichlet energy on the graph, subject to a Dirichlet condition $u = g$ on Γ. Minimizers u are harmonic functions on the graph, and thus the problem can be view as harmonic extension.

It has been observed [18,34] that when the size of Γ (the labeled points) is small, the performance of graph Laplacian learning algorithms degrades substantially. In practice, the learned function u fails to attain the conditions $u = g$ on Γ continuously, and degenerates into a constant label function that provides little information about the machine learning problem. Figure 1 gives an example of this issue. There are several ways to explain this degeneracy. First, in the limit of infinite data, the variational problem (1) is consistent with the continuum Dirichlet problem

$$\min_u \int_\Omega |\nabla u|^2 \, dx, \tag{2}$$

subject to a boundary condition $u = g$ on $\Gamma \subset \Omega \subset \mathbb{R}^d$. If Γ is finite this problem is ill-posed since the trace of an $H^1(\Omega)$ function at a point is not well-defined. In particular, there are minimizing sequences for the constrained problem converging to a constant function outside of Γ for which the Dirichlet energy converges to zero. In particular the minimum is not attained. From another perspective, minimizers of the continuum Dirichlet problem (2) satisfy Laplace's equation $\Delta u = 0$ with Dirichlet condition $u = g$ on Γ, and Laplace's equation is not well-posed without some boundary regularity (an exterior sphere condition), which does not hold for isolated points. In both cases, we are simply observing that the capacity of a point is zero in dimensions $d \geqslant 2$.

Several methods have been proposed recently to address the degeneracy of Laplacian learning with few labels. In [18], a class of p-Laplacian learning algorithms was proposed, which replace the exponent 2 in (1) with $p > 2$. The p-Laplacian models were considered previously for other applications [2,9,18,50], and the $p \to \infty$ case, which is called Lipschitz learning, was considered in [29,33]. The idea behind the p-Laplacian models is that the continuum variational problem is now the p-Dirichlet problem

$$\min_u \int_\Omega |\nabla u|^p \, dx, \tag{3}$$

and for $p > d$ the Sobolev embedding $W^{1,p}(\Omega) \hookrightarrow C^{0,\alpha}(\overline{\Omega})$ allows the assignment of boundary values at isolated points. The p-Laplacian models, including the $p = \infty$ version, were proven to be well-posed in the limit of infinite unlabeled data and finite labeled data precisely when $p > d$ in [10,11,41]. The disadvantage of p-Laplacian models is that the nonlinearity renders them more computationally challenging to solve, compared with standard Laplacian regularization. Other approaches include higher order Laplacian regularization [6,17,49] and using a spectral cut-off [5].

The approach most closely related to our work is the *weighted nonlocal Laplacian* of Shi et al. [38], which replaces the learning problem (1) with

$$\min_{u:\mathcal{X}\to\mathbb{R}} \sum_{x\in\mathcal{X}\setminus\Gamma} \sum_{y\in\mathcal{X}} w_{xy}(u(x) - u(y))^2 + \mu \sum_{x\in\Gamma} \sum_{y\in\mathcal{X}} w_{xy}(g(x) - u(y))^2, \tag{4}$$

where $\mu > 0$ is selected as the ratio of unlabeled to labeled data. The method increases the weights of edges adjacent to labels, which encourages the label function to be flat near labels. The authors show in [38] that the method produces superior results, compared to the standard graph Laplacian, for classification with very few labels. Furthermore, since the method is a standard graph Laplacian with a modified weight matrix, it has similar computational complexity to Laplacian learning, and is fast compared to the non-linear p-Laplace methods, for example. However, as we prove in this paper, the weighted nonlocal Laplacian of [38] becomes ill-posed (degenerate) in the limit of infinite unlabeled and finite labeled data. This is a direct consequence of Corollary 3.8. Numerical simulations in Sect. 5 illustrate the way in which the method becomes degenerate. The issue is the same as for Laplacian learning, since the weights

are modified only locally near label points and the size of this neighborhood shrinks to zero in the large sample size limit.

1.1 Properly Weighted Laplacian

In this paper, we show how to properly weight the graph Laplacian so that it remains well-posed in the limit of infinite unlabeled and finite labeled data. Our method, roughly speaking, modifies the problem to one of the form:

$$\begin{aligned} &\text{Minimize} \quad \sum_{x,y\in\mathcal{X}} \gamma(x) w_{xy} (u(x)-u(y))^2 \quad \text{over } u : \mathcal{X} \to \mathbb{R}, \\ &\text{subject to constraint} \quad u(x) = g(x) \text{ for all } x \in \Gamma \end{aligned} \tag{5}$$

where $\gamma(x) = \text{dist}(x, \Gamma)^{-\alpha}$ and $\alpha > d-2$ (see Sect. 1.2 for precise definitions). Here, we are modifying the weights not just of edges connecting to points of Γ, but also in a neighborhood of Γ. We show that this model is stable as the number unlabeled data points increases to infinity, under appropriate scaling of the graph construction. In particular we show that the minimizers of the graph problem above converge as the number of unlabeled data points increases to the minimizer of a "continuum learning problem". We give the precise assumptions on the discrete model below and describe the continuum problem in Sect. 2. Here we give a brief explanation as to why $\alpha > d-2$ is the natural scaling for the weight.

To illustrate what is happening near a labeled point, consider $\Gamma = \{0\}$ and take the domain from which the points are sampled to be the unit ball $\Omega = B(0, 1)$ in $\mathbb{R}^d$. The continuum variational problem corresponding to (5) involves minimizing

$$I[u] = \int_{B(0,1)} |x|^{-\alpha} |\nabla u|^2 \, dx. \tag{6}$$

The Euler–Lagrange equation satisfied by minimizers of I is

$$\text{div}\left(|x|^{-\alpha} \nabla u\right) = 0. \tag{7}$$

This equation has a radial solution $u(x) = |x|^{\alpha+2-d}$, which is continuous at $x = 0$ when $\alpha > d-2$. This suggests the solutions will assume this radial profile near labels, and the model will be well-posed for $\alpha > d-2$. Furthermore when $\alpha \geqslant d-1$ one can expect the solution to be Lipschitz near labels, and for $\alpha \geqslant d$ it is should be differentiable at the labels. It is important to point out that the proper weighting changes the degenerate limiting continuum problem to one that is well-posed with "boundary" data at isolated points.

We now provide a precise description of the properly-weighted graph Laplacian.

1.2 Model and Definitions

Let $\Omega \subset \mathbb{R}^d$ be open and bounded with a Lipschitz boundary. Let $\Gamma \subset \Omega$ be a finite collection of points along with a given label function $g : \Gamma \to \mathbb{R}$. Let μ be a probability measure on Ω with continuous density ρ which is bounded from above and below by positive constants. Let $x_1, x_2, \ldots, x_n$ be independent and identically distributed random variables with distribution μ, and let

$$X_n := \{x_1, x_2, \ldots, x_n\},$$

and $\mathcal{X}_n := X_n \cup \Gamma$. To define the edge weights we use a radial kernel η with profile $\boldsymbol{\eta} : [0, \infty) \to [0, \infty)$ which is nonincreasing, continuous at 0 and satisfies

$$\begin{cases} \boldsymbol{\eta}(t) \geqslant 1, & \text{if } 0 \leqslant t \leqslant 1 \\ \boldsymbol{\eta}(t) = 0, & \text{if } t > 2. \end{cases} \tag{8}$$

All of the results we state can be extended to kernels which decay sufficiently fast, in particular the Gaussian. For $\varepsilon > 0$ we define the rescaled kernel

$$\eta_\varepsilon(x - y) = \frac{1}{\varepsilon^d} \boldsymbol{\eta}\left(\frac{|x - y|}{\varepsilon}\right). \tag{9}$$

We now introduce the penalization of the gradient, which is heavier near labeled points. Let $R > 0$ be the minimum distance between pairs of points in Γ:

$$R = \min\{|x - y| \ : \ x, y \in \Gamma, \ \ x \neq y\}. \tag{10}$$

For $r_0 > 0$ and $\alpha \geqslant 0$ let $\gamma \in C^\infty(\overline{\Omega} \backslash \Gamma)$ be any function satisfying $\gamma \geqslant 1$ on Ω and

$$\gamma(x) = 1 + \left(\frac{r_0}{\operatorname{dist}(x, \Gamma)}\right)^\alpha \text{ whenever } \operatorname{dist}(x, \Gamma) \leqslant \frac{R}{4}, \tag{11}$$

where $\operatorname{dist}(x, \Gamma)$ denotes the Euclidean distance from x to the closest point in Γ. For $\zeta > 1$ we set

$$\gamma_\zeta(x) = \min\{\gamma(x), \zeta\}. \tag{12}$$

For $u \in L^2(\mathcal{X}_n)$ we define the energy

$$\mathcal{GE}_{n,\varepsilon,\zeta}(u) = \frac{1}{n^2 \varepsilon^2} \sum_{x,y \in \mathcal{X}_n} \gamma_\zeta(x) \eta_\varepsilon(x - y) |u(x) - u(y)|^2. \tag{13}$$

The Laplacian learning problem is to

$$\text{minimize } \ \mathcal{GE}_{n,\varepsilon,\zeta}(u) \ \text{ over } \left\{u \in L^2(\mathcal{X}_n) \text{ and } u = g \text{ on } \Gamma\right\}. \tag{14}$$

We note that the unique minimizer $u \in L^2(\mathcal{X}_n)$ of (14) satisfies the optimality condition

$$\begin{cases} \mathcal{GL}_{n,\varepsilon,\zeta} u(x) &= 0, \quad \text{if } x \in X_n \ (= \mathcal{X}_n \backslash \Gamma) \\ u(x) &= g(x), \quad \text{if } x \in \Gamma, \end{cases} \tag{15}$$

where $\mathcal{GL}_{n,\varepsilon,\zeta} : L^2(\mathcal{X}_n) \to L^2(\mathcal{X}_n)$ is the graph Laplacian, given by

$$\mathcal{GL}_{n,\varepsilon,\zeta} u(x) = \frac{1}{2n\varepsilon^2} \sum_{y \in \mathcal{X}_n} (\gamma_\zeta(x) + \gamma_\zeta(y)) \eta_\varepsilon(x-y)(u(y) - u(x)). \tag{16}$$

Some remarks about the model are in order.

Remark 1.1 When considering the discrete functional, ζ depends on n and diverges to infinity (sufficiently fast) as $n \to \infty$. The constant r_0 represents the length scale of the crossover from the strong local penalization near Γ to uniform far-field penalization. The introduction of ζ is needed since $\gamma(x) = \infty$ on Γ and so using γ directly would impose a hard constraint on neighbors of labeled points. While we can allow $\zeta = \infty$ in our model by interpreting products $\infty \cdot 0$ as 0, we wanted to allow for a model with far less stringent constraints on agreement with the labeled points in the immediate vicinity of Γ. We note that the critical distance to Γ, when γ_ζ crosses over from γ to ζ equals

$$r_\zeta = r_0(\zeta - 1)^{-1/\alpha} \quad \text{provided that } r_0(\zeta-1)^{-1/\alpha} < \frac{R}{4}. \tag{17}$$

Remark 1.2 In practice, one can take (11) to be the definition of the weights on the whole domain Ω. We only need γ to be smooth for a part of our analysis in Sect. 2.2. The issue is that since the distance function $d(x, \Gamma)$ is not differentiable (it is only Lipschitz on $\Omega \backslash \Gamma$ if Γ has more than one point), γ cannot be both smooth and globally given by (11). To elaborate, γ appears as part of the diffusion coefficient in the limiting elliptic problem (see Eq. (22)). The solutions have nicer regularity properties when we take γ to be smooth, away from the labels. For the other results we only need that γ is bounded from below by a positive number and has singularities, with a particular growth rate, near the points of Γ.

Remark 1.3 Instead of truncating γ at the radius r_ζ to construct the weights γ_ζ, we can take a possibly discontinuous model of the form

$$\gamma_{\zeta,r}(x) = \begin{cases} \gamma(x), & \text{if } \operatorname{dist}(x, \Gamma) > r \\ \zeta, & \text{if } \operatorname{dist}(x, \Gamma) \leqslant r. \end{cases} \tag{18}$$

This model is more general, since we can set $r = r_\zeta$ to recover (12). Choosing $\zeta \gg 1 + (r_0/r)^\alpha$ places a larger penalty on the gradient in the inner region where $\operatorname{dist}(x, \Gamma) \leqslant r$, compared to Eq. (12). This model is useful in the analysis of the graph

based problem, and gives a sharper result for continuity at the labels (see Remark 4.3). In the limit as $n \to \infty$ we would take $\zeta \to \infty$ and $r \to 0$ with $r \geqslant r_\zeta$.

Remark 1.4 We remark that the discrete functional (13) can be rewritten as

$$\mathcal{GE}_{n,\varepsilon,\zeta}(u) = \frac{1}{2n^2\varepsilon^2} \sum_{x,y \in \mathcal{X}_n} (\gamma_\zeta(x) + \gamma_\zeta(y))\eta_\varepsilon(x-y)|u(x) - u(y)|^2, \tag{19}$$

and so the problem has a symmetric weight matrix.

1.3 Outline

The continuum properly-weighted Dirichlet energy, which describes the asymptotic behavior of the properly-weighted graph Laplacian (14) is presented in Sect. 2 (Eqs. (20) and (21)). To show that the continuum problem is well posed and to establish its basic properties, in Sect. 2 we also study properties of singularly weighted Sobolev spaces. In particular the Trace Theorem 2.2 plays a key role in showing that the data can be imposed on a set of isolated points, which enables us to show the well-posedness in Theorem 2.7. The Euler–Lagrange equation of the variational problem is the elliptic problem we study in Sect. 2.2. In particular we show that solutions are C^2 away from the labels and Hölder continuous globaly.

In Sect. 3 we turn to asymptotics of the graph-based problems. We prove in Theorem 3.1 that the solutions of the graph-based learning problem (14), for the properly-weighted Laplacian, converge in the large sample size limit to the solution of a continuum variational problem (20)–(21). We achieve this by showing the Γ-convergence of the discrete variational problems to the corresponding continuum problem. We also prove a negative result, showing that the nonlocal weighted Laplacian [38] is degenerate (ill-posed) in the large data limit (with fixed number of labeled points). In Sect. 4.1 we prove that solutions of the graph-based learning problem for the properly-weighted Laplacian attain their labeled values continuously with high probability (Theorem 4.1). In Sect. 5 we present the results of numerical simulations illustrating the estimators obtained by our method, and its performance in classification tasks on synthetic data and in classifying handwritten digits from the MNIST dataset [30]. The classification problems on synthetic data contrast the stability of the properly-weighted Laplacian with the instability of the standard graph Laplacian and related methods. The MNIST experiments show superior performance of our method compared to the standard graph Laplacian, and similar performance to the weighted Laplacian of [38]. In the Appendix A we recall some background results used and show and auxiliary technical result.

2 Analysis of the Continuum Problem

The continuum variational problem corresponding to the graph-based problem (14) is

$$\text{minimize } \mathcal{E}(u) \text{ over } \left\{ u \in H^1_\gamma(\Omega) \text{ and } u = g \text{ on } \Gamma \right\}, \tag{20}$$

where $\mathcal{E}$ is given by

$$\mathcal{E}(u) = \frac{1}{2} \int_\Omega \gamma |\nabla u|^2 \rho^2 \, dx, \tag{21}$$

ρ is continuous and bounded from above and below by positive constants, and the weighted Sobolev Space $H^1_\gamma(\Omega)$ is defined by (25). It follows from Lemma 2.1 that for γ which grow near points of Γ as fast as or faster than $\text{dist}(x, \Gamma)^{-\alpha}$, the functions in $H^1_\gamma(\Omega)$ have a trace at Γ (defined by (33)), which enables one to assign the condition $u = g$ on Γ in (20).

The Euler–Lagrange equation satisfied by minimizers of (21) is the elliptic equation

$$\begin{cases} -\operatorname{div}(\gamma \rho^2 \nabla u) = 0 & \text{in } \Omega \backslash \Gamma \\ u = g & \text{on } \Gamma \\ \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial\Omega. \end{cases} \tag{22}$$

In this section we study the variational problem (20) and the elliptic problem (22) rigorously. The theory is nonstandard due to the boundary condition $u = g$ on Γ, since Γ is a collection of isolated points and does not satisfy an exterior sphere condition. As a consequence of this analysis, we prove in Sect. 4.1 that solutions of the graph-based problem are continuous at the labels.

Before studying this problem, we need to perform a careful analysis of a particular weighted Sobolev space.

2.1 Weighted Sobolev Spaces

In this section we study the Sobolev space with norm weighted by γ. While there exists a rich literature on Weighted Sobolev Spaces, we did not find the precise results we need. Below we develop a self-contained, but brief, description of the spaces with particular weights of interest.

For $u \in H^1(\Omega)$ we define

$$[u]^2_{H^1_\gamma(\Omega)} = \int_\Omega \gamma |\nabla u|^2 \, dx, \tag{23}$$

and

$$\|u\|^2_{H^1_\gamma(\Omega)} = \|u\|^2_{L^2(\Omega)} + [u]^2_{H^1_\gamma(\Omega)}. \tag{24}$$

We define

$$H^1_\gamma(\Omega) = \left\{ u \in H^1(\Omega) \, : \, \|u\|_{H^1_\gamma(\Omega)} < \infty \right\}, \tag{25}$$

and endow $H^1_\gamma(\Omega)$ with the norm $\|u\|_{H^1_\gamma(\Omega)}$. We also denote by $H^1_{\gamma,0}(\Omega)$ the closure of $C^\infty_c(\overline{\Omega}\backslash\Gamma)$ in $H^1_\gamma(\Omega)$. The space $H^1_\gamma(\Omega)$ is the natural function space on which to pose the variational problem (20).

Throughout this section we let B_r denote the open ball of radius $r > 0$ centered at the origin in $\mathbb{R}^d$. Whenever we consider the space $H^1_\gamma(B_r)$, we will implicitly assume the choice of $\gamma(x) = |x|^{-\alpha}$. Hence

$$[u]^2_{H^1_\gamma(B_r)} = \int_{B_r} |\nabla u|^2 |x|^{-\alpha}\, dx. \tag{26}$$

In all other occurrences, γ is defined as in Sect. 1.2, and in particular we always assume (11) holds. We also use the notation $(u)_{x,r} = \fint_{B(x,r)} u\, dx$ for the average of u over the ball $B(x, r)$, and $(u)_r := (u)_{0,r}$. We also assume in this section that Ω has a Lipschitz boundary.

First, we study the trace of $H^1_\gamma(\Omega)$ functions on Γ. Before proving a general trace theorem, we require a preliminary lemma.

Lemma 2.1 *Let $\alpha > d - 2$ and $u \in H^1_\gamma(B_r)$. Then $x = 0$ is a Lebesgue point* [35] *for u, i.e., $u(0) := \lim_{\varepsilon\to 0}(u)_\varepsilon$ exists, and*

$$|u(0) - (u)_\varepsilon| \leqslant C\varepsilon^{(\alpha+2-d)/2}[u]_{H^1_\gamma(B_\varepsilon)} \tag{27}$$

for all $0 < \varepsilon \leqslant r$.

Proof We compute

$$\int_{B_\varepsilon} |\nabla u|^2\, dx \leqslant \int_{B_\varepsilon} |\nabla u|^2 |x|^{-\alpha} \varepsilon^\alpha\, dx \leqslant [u]^2_{H^1_\gamma(B_\varepsilon)} \varepsilon^\alpha.$$

By the Poincaré inequality we have

$$\fint_{B_\varepsilon} (u - (u)_\varepsilon)^2\, dx \leqslant Cr^2 \fint_{B_\varepsilon} |\nabla u|^2\, dx \leqslant C[u]^2_{H^1_\gamma(B_\varepsilon)} \varepsilon^{\alpha+2-d}. \tag{28}$$

For $0 < s < t \leqslant r$ we apply (28) with s and t in place of ε to obtain

$$\begin{aligned} s^d((u)_s - (u)_t)^2 &\leqslant C \int_{B_s} ((u)_s - (u)_t)^2\, dx \\ &\leqslant C \int_{B_s} ((u)_s - u)^2\, dx + C \int_{B_s} ((u)_t - u)^2\, dx \\ &\leqslant C[u]^2_{H^1_\gamma(B_t)}(s^{\alpha+2} + t^{\alpha+2}). \end{aligned} \tag{29}$$

For $0 < q < \varepsilon \leqslant r$ with $\varepsilon \leqslant 4q$ we can set $s = q$ and $t = \varepsilon$ above to obtain

$$|(u)_q - (u)_\varepsilon|^2 \leqslant C[u]^2_{H^1_\gamma(B_\varepsilon)} \varepsilon^{\alpha+2-d}. \tag{30}$$

For $0 < q < \varepsilon \leqslant r$ with $\varepsilon > 4q$, let $k \in \mathbb{N}$ be the greatest integer smaller than $\log_2(\varepsilon/q)$. Since $\varepsilon > 2q$, we have $k \geqslant 1$. Choose $b > 0$ so that $b^k = \varepsilon/q$. Then

$$\log(b) = \frac{\log(\varepsilon/q)}{k} \geqslant \frac{\log(\varepsilon/q)}{\log_2(\varepsilon/q)} \geqslant \log(2)$$

and

$$\log(b) \leqslant \frac{\log(\varepsilon/q)}{\log_2(\varepsilon/q) - 1} = \frac{\log(2)\log_2(\varepsilon/q)}{\log_2(\varepsilon/q) - 1} \leqslant 2\log(2),$$

since $\log_2(\varepsilon/q) > 2$. Therefore $2 \leqslant b \leqslant 4$. Let us set $\varepsilon_j = \varepsilon b^{-j}$ and $a_j = (u)_{\varepsilon_j}$. Then $\varepsilon_0 = \varepsilon$ and $\varepsilon_k = \varepsilon b^{-k} = q$. Setting $t = \varepsilon_j$ and $s = \varepsilon_{j+1}$ in (29) yields

$$|a_j - a_{j+1}|^2 \leqslant C[u]^2_{H^1_\gamma(B_\varepsilon)}(\varepsilon_{j+1}^{\alpha+2-d} + \varepsilon_j^{\alpha+2}\varepsilon_{j+1}^{-d}) \leqslant C[u]^2_{H^1_\gamma(B_\varepsilon)} b^{-j(\alpha+2-d)}\varepsilon^{\alpha+2-d}.$$

Therefore

$$|(u)_q - (u)_\varepsilon| \leqslant \sum_{j=0}^{k-1} |a_{j+1} - a_j| \leqslant C[u]_{H^1_\gamma(B_\varepsilon)}\varepsilon^{(\alpha+2-d)/2} \tag{31}$$

holds for all $k \geqslant 1$, where C is independent of u, ε and k.

In either case, we have established that

$$|(u)_q - (u)_\varepsilon|^2 \leqslant C[u]^2_{H^1_\gamma(B_\varepsilon)}\varepsilon^{\alpha+2-d} \tag{32}$$

holds for all $0 < q < \varepsilon \leqslant r$. Thus, the sequence $\varepsilon \mapsto (u)_\varepsilon$ is Cauchy and converges to a real number as $\varepsilon \to 0$. Sending $q \to 0$ in (32) completes the proof. □

By Lemma 2.1, we can define the trace operator $\mathrm{Tr} : H^1_\gamma(\Omega) \to \mathbb{R}^\Gamma$ by

$$\mathrm{Tr}[u](x) = \lim_{\varepsilon \to 0} \fint_{B(x,\varepsilon)} u\,dx \quad (x \in \Gamma). \tag{33}$$

We endow $\mathbb{R}^\Gamma$ with the Euclidean norm. We now prove our main trace theorem.

Theorem 2.2 (Trace Theorem) *Let $\alpha > d - 2$ and assume γ satisfies (11) and Ω has a Lipschitz boundary. Then the trace operator $\mathrm{Tr} : H^1_\gamma(\Omega) \to \mathbb{R}^\Gamma$ is bounded, and satisfies $\mathrm{Tr}[u](x) = u(x)$ whenever u is continuous at $x \in \Gamma$. Furthermore, for every $u, v \in H^1_\gamma(\Omega)$ with $\|u - v\|^{2/(\alpha+2)}_{L^2(\Omega)} \leqslant R/2$ we have*

$$|\mathrm{Tr}[u] - \mathrm{Tr}[v]| \leqslant C(1 + [u]_{H^1_\gamma(\Omega)} + [v]_{H^1_\gamma(\Omega)})\|u - v\|^{1-d/(\alpha+2)}_{L^2(\Omega)}. \tag{34}$$

Proof By Lemma 2.1 each $x \in \Gamma$ is a Lebesgue point of u, and we have for $r \leqslant R/2$

$$\begin{aligned} |\operatorname{Tr}[u](x)| &\leqslant \fint_{B(x,r)} u\,dx + Cr^{(\alpha+2-d)/2}[u]_{H^1_\gamma(\Omega)} \\ &\leqslant Cr^{-d/2}\|u\|_{L^2(\Omega)} + Cr^{(\alpha+2-d)/2}\|u\|_{H^1_\gamma(\Omega)}, \end{aligned}$$

where $R > 0$ is defined in (10). Fixing $r = R/2$ we have $|\operatorname{Tr}[u](x)| \leqslant C\|u\|_{H^1_\gamma(\Omega)}$, hence $\operatorname{Tr} : H^1_\gamma(\Omega) \to \mathbb{R}$ is bounded.

To prove (34), let $u, v \in H^1_\gamma(\Omega)$ and $x \in \Gamma$. For simplicity, we write $u(x)$ and $v(x)$ for $\operatorname{Tr}[u](x)$ and $\operatorname{Tr}[v](x)$, respectively. By Lemma 2.1 we have for $0 < \varepsilon \leqslant R/2$

$$\begin{aligned} |u(x) - v(x)| &\leqslant \big|u(x) - (u)_{x,\varepsilon}\big| + \big|v(x) - (v)_{x,\varepsilon}\big| + \big|(u)_{x,\varepsilon} - (v)_{x,\varepsilon}\big| \\ &\leqslant C\varepsilon^{(\alpha+2-d)/2}([u]_{H^1_\gamma(B(x,\varepsilon))} + [v]_{H^1_\gamma(B(x,\varepsilon))}) + C\varepsilon^{-d}\int_{B_\varepsilon} |u - v|\,dx \\ &\leqslant C\varepsilon^{(\alpha+2-d)/2}([u]_{H^1_\gamma(\Omega)} + [v]_{H^1_\gamma(\Omega)}) + C\varepsilon^{-d/2}\|u - v\|_{L^2(\Omega)}. \end{aligned}$$

Choosing $\varepsilon = \|u - v\|^{2/(\alpha+2)}_{L^2(\Omega)}$, we obtain

$$|u(x) - v(x)| \leqslant C(1 + [u]_{H^1_\gamma(\Omega)} + [v]_{H^1_\gamma(\Omega)})\|u - v\|^{1-d/(\alpha+2)}_{L^2(\Omega)},$$

provided $\varepsilon \leqslant R/2$. □

We now examine the decay of the L^2 norm of trace zero functions.

Lemma 2.3 *Let $\alpha > d - 2$ and $u \in H^1_\gamma(B_r)$ with $\operatorname{Tr}[u](0) = 0$. Then*

$$\fint_{\partial B_\varepsilon} u^2\,dS + \fint_{B_\varepsilon} u^2\,dx \leqslant C\varepsilon^{\alpha+2-d}[u]^2_{H^1_\gamma(B_\varepsilon)} \tag{35}$$

for all $0 < \varepsilon \leqslant r$.

Proof Since $\operatorname{Tr}[u](0) = 0$, Lemma 2.1 yields

$$\|(u)_\varepsilon\|_{L^2(B_\varepsilon)} \leqslant C\varepsilon^{d/2}|(u)_\varepsilon| \leqslant C\varepsilon^{(\alpha+2)/2}[u]_{H^1_\gamma(B_\varepsilon)}.$$

Recalling (28) from the proof of Lemma 2.1 we deduce

$$\|u - (u)_\varepsilon\|_{L^2(B_\varepsilon)} \leqslant C\varepsilon^{(\alpha+2)/2}[u]_{H^1_\gamma(B_\varepsilon)}.$$

Therefore

$$\fint_{B_\varepsilon} u^2\,dx = C\varepsilon^{-d}\|u\|^2_{L^2(B_\varepsilon)} \leqslant C\varepsilon^{\alpha+2-d}[u]^2_{H^1_\gamma(B_\varepsilon)},$$

which establishes one part of (35).

For the other part, we use a standard trace estimate that we include for completeness. We have

$$\begin{aligned}\varepsilon \int_{\partial B_\varepsilon} u^2\, dS &= \int_{B_\varepsilon} \operatorname{div}(xu^2)\, dx \\ &= \int_{B_\varepsilon} du^2 + 2u\nabla u \cdot x\, dx \\ &\leqslant C\int_{B_\varepsilon} u^2\, dx + C\varepsilon^2 \int_{B_\varepsilon} |\nabla u|^2\, dx \\ &\leqslant C\int_{B_\varepsilon} u^2\, dx + C\varepsilon^{\alpha+2}[u]^2_{H^1_\gamma(B_\varepsilon)}.\end{aligned}$$

Dividing both sides by ε^d we obtain

$$⨍_{\partial B_\varepsilon} u^2\, dS \leqslant C⨍_{B_\varepsilon} u^2\, dx + C\varepsilon^{\alpha+2-d}[u]^2_{H^1_\gamma(B_\varepsilon)},$$

which completes the proof. □

We now show that trace zero functions can be approximated in $H^1_\gamma(\Omega)$ by smooth functions compactly supported away from Γ.

Theorem 2.4 (Trace zero functions) *Let $\alpha > d-2$ and assume γ satisfies (11) and Ω has a Lipschitz boundary. Then $u \in H^1_{\gamma,0}(\Omega)$ if and only if $u \in H^1_\gamma(\Omega)$ and $\operatorname{Tr}[u] = 0$.*

Proof If $u \in H^1_{\gamma,0}(\Omega)$, then there exists $u_k \in C^\infty_c(\overline{\Omega}\backslash\Gamma)$ so that $u_k \to u$ in $H^1_\gamma(\Omega)$. In particular, u_k is uniformly bounded in $H^1_\gamma(\Omega)$. Thus, by Theorem 2.2, we have $\operatorname{Tr}[u](x) = \lim_{k\to\infty} u_k(x) = 0$ for each $x \in \Gamma$.

Conversely, let $u \in H^1_\gamma(\Omega)$ such that $\operatorname{Tr}[u] = 0$. Without loss of generality, we may assume $\Omega = B_r$, $\Gamma = \{0\}$, and $\operatorname{Tr}[u](0) = 0$. Choose a smooth nonincreasing function $\xi : [0,\infty) \to [0,1]$ such that $\xi(t) = 1$ for $0 \leqslant t \leqslant 1$ and $\xi(t) = 0$ for $t \geqslant 2$. For a positive integer $k \geqslant 1$ define $\xi_k(x) = \xi(k|x|)$ and $w_k = u(1-\xi_k)$. We compute

$$\begin{aligned}\int_{B_r} |\nabla w_k - \nabla u|^2 |x|^{-\alpha}\, dx &= \int_{B_r} \left|\xi_k \nabla u + ku\xi'(k|x|)\frac{x}{|x|}\right|^2 |x|^{-\alpha}\, dx \\ &\leqslant C\int_{B_{2/k}} |\nabla u|^2|x|^{-\alpha}\, dx + Ck^2 \int_{B_{2/k}\backslash B_{1/k}} u^2|x|^{-\alpha}\, dx \\ &\leqslant C\int_{B_{2/k}} |\nabla u|^2|x|^{-\alpha}\, dx + Ck^{\alpha+2}\int_{B_{2/k}} u^2\, dx \\ &\leqslant C\int_{B_{2/k}} |\nabla u|^2|x|^{-\alpha}\, dx, \end{aligned} \tag{36}$$

the last line following from Lemma 2.3. Therefore $w_k \to u$ in $H^1_\gamma(B_r)$ as $k \to \infty$. To produce a smooth approximating sequence u_k, we simply mollify the sequence w_k. □

As a corollary, we can prove density of smooth functions that are locally constant near Γ.

Corollary 2.5 *For any $\alpha > 0$ the set*

$$\mathcal{S} = \{u \in C^\infty(\overline{\Omega}) \ : \ (\exists s > 0)(\forall x \in \Gamma)(\forall z \in B(0,1)) \ \ u(x + sz) = u(x)\}$$

is a dense subset of $H^1_\gamma(\Omega)$.

Proof We split the proof into two cases.

Case 1: $\alpha > d - 2$. Let $u \in H^1_\gamma(\Omega)$. There exists $\psi \in \mathcal{S}$ such that $\mathrm{Tr}[\psi] = \mathrm{Tr}[u]$. Since $w := u - \psi \in H^1_{\gamma,0}(\Omega)$, there exists by Theorem 2.4 a sequence $\varphi_k \in C^\infty_c(\overline{\Omega}\backslash\Gamma)$ such that $\varphi_k \to w$ as $k \to \infty$. We simply note that $\psi_k := \varphi_k + \psi \in \mathcal{S}$ and $\psi_k \to u$ in $H^1_\gamma(\Omega)$ as $k \to \infty$.

Case 2: $\alpha \leqslant d-2$. In this case, $C^\infty(\overline{\Omega})$ is dense in $H^1_\gamma(\Omega)$ by a standard mollification argument, since the weighting kernel $|x|^{-\alpha}$ is integrable. Hence, for $u \in H^1_\gamma(\Omega)$ with $\alpha \leqslant d-2$ there exists $\varphi_k \in C^\infty(\overline{\Omega})$ such that $\varphi_k \to u$ in $H^1_\gamma(\Omega)$. Since φ_k is smooth, we automatically have

$$\int_\Omega \mathrm{dist}(x, \Gamma)^{1-d} |\nabla \varphi_k|^2 \, dx < \infty.$$

Thus, by case 1, there exists a sequence $\psi_{k,j} \in \mathcal{S}$ such that for each k, $\psi_{k,j} \to \varphi_k$ in H^1_γ as $j \to \infty$, since $\alpha \leqslant d - 2 \leqslant d - 1$. The proof is completed with a diagonal argument. □

Finally, we prove a Hardy-type inequality for trace zero functions in $H^1_\gamma(B_R)$.

Theorem 2.6 (Hardy's inequality) *Let $\alpha > d - 2$ and assume γ satisfies (11) and Ω has a Lipschitz boundary. If $u \in H^1_\gamma(B_r)$ with $\mathrm{Tr}[u](0) = 0$ then $\frac{u}{|x|^{(\alpha+2)/2}} \in L^2(B_r)$ and*

$$\int_{B_r} \frac{u^2}{|x|^{\alpha+2}} \, dx \leqslant C[u]^2_{H^1_\gamma(B_r)}. \tag{37}$$

Proof By a change of variables we can reduce to the case of $r = 1$. We first note that

$$\mathrm{div}\left(\frac{x}{|x|^{\alpha+2}}\right) = -\frac{\alpha + 2 - d}{|x|^{\alpha+2}}$$

for $x \neq 0$. Thus, for $\varepsilon > 0$ we have

$$\begin{aligned}\int_{B_1\backslash B_\varepsilon} \frac{u^2}{|x|^{\alpha+2}} \, dx &= -\frac{1}{\alpha + 2 - d} \int_{B_1\backslash B_\varepsilon} u^2 \, \mathrm{div}\left(\frac{x}{|x|^{\alpha+2}}\right) dx \\ &= \frac{1}{\alpha + 2 - d}\left[2 \int_{B_1\backslash B_\varepsilon} u \nabla u \cdot \frac{x}{|x|^{\alpha+2}} \, dx\right.\end{aligned}$$

$$
\begin{aligned}
&\quad - \int_{\partial B_1} u^2 \, dS + \frac{1}{\varepsilon^{\alpha+1}} \int_{\partial B_\varepsilon} u^2 \, dS \Bigg] \\
&\leqslant C \int_{B_1 \setminus B_\varepsilon} |u||\nabla u||x|^{-\alpha-1} \, dx + C[u]^2_{H^1_\gamma(B_1)} \\
&\quad + C \int_{B_\varepsilon} |x|^{-\alpha} |\nabla u|^2 \, dx,
\end{aligned}
$$

where the last line follows from Lemma 2.3 and the assumption $\alpha > d - 2$. Applying Cauchy's inequality to the first term and rearranging yields

$$
\int_{B_1 \setminus B_\varepsilon} \frac{u^2}{|x|^{\alpha+2}} \, dx \leqslant C[u]^2_{H^1_\gamma(B_1)} + C \int_{B_\varepsilon} |x|^{-\alpha} |\nabla u|^2 \, dx.
$$

Sending $\varepsilon \to 0$ completes the proof. □

We now establish the well posedness of the continuum properly-weighted Laplacian learning problem.

Theorem 2.7 *Let $\alpha > d - 2$, and assume ρ is continuous and bounded above and below by positive constants, γ satisfies (11), and Ω has a Lipschitz boundary. Then the problem (20) has a unique solution.*

Proof The existence follows by the direct method of the calculus of variations. Namely let u_k, $k = 1, 2, \ldots$ be a minimizing sequence. By the Sobolev Embedding Theorem, u_k has a subsequence which converges weakly in $H^1_\gamma(\Omega)$ and in $L^2(\Omega)$ towards $u \in H^1_\gamma(\Omega)$. Since $\mathcal{E}$ is convex, it is weakly lower-semicontinuous and thus $\mathcal{E}(u) \leqslant \liminf_{k\to\infty} \mathcal{E}(u_k)$. Furthermore note that (34) implies that $\mathrm{Tr}(u_k)(z) \to \mathrm{Tr}(u)(z)$ for every $z \in \Gamma$. Thus $u = g$ on Γ. We conclude that u is the desired minimizer. The uniqueness follows from convexity of $\mathcal{E}$, by a standard argument, which is recalled in the proof of Lemma 2.11. □

2.2 Elliptic Problem

We now study the elliptic Euler–Lagrange equation (22). We additionally assume in this section that Ω has a $C^{2,\alpha}$ boundary and $\rho \in C^{1,\sigma}(\overline{\Omega})$ for some $\sigma > 0$. As before, we assume ρ is bounded above and below by positive constants.

Definition 2.8 We say that $u \in H^1_\gamma(\Omega)$ is a weak solution of (22) if

$$
\int_\Omega \gamma \rho^2 \nabla u \cdot \nabla \varphi \, dx = 0 \tag{38}
$$

for all $\varphi \in H^1_{\gamma,0}(\Omega)$ and $\mathrm{Tr}[u](x) = g(x)$ for all $x \in \Gamma$.

We first need a preliminary proposition on barrier functions.

Proposition 2.9 (Barrier) *Let $\alpha > d-2$ and fix any $0 < \beta < \alpha + 2 - d$. Then there exists $c > 0$ depending on α, β, ρ and d such that $w(x) = |x|^\beta$ satisfies*

$$\operatorname{div}(\rho^2(1+|x|^{-\alpha})\nabla w) \leqslant -\frac{\beta}{2}(\alpha+2-\beta-d)\rho^2|x|^{-(\alpha+2-\beta)} \tag{39}$$

for all $0 < |x| \leqslant c$.

Proof. Since $\nabla w(x) = \beta|x|^{\beta-2}x$ we have

$$\begin{aligned}
&\operatorname{div}(\rho^2(1+|x|^{-\alpha})\nabla w)\\
&\quad = \beta \operatorname{div}(\rho^2|x|^{\beta-2}x) + \beta \operatorname{div}(\rho^2|x|^{\beta-\alpha-2}x)\\
&\quad = 2\beta|x|^{\beta-\alpha-2}(1+|x|^\alpha)\rho\nabla\rho\cdot x + \beta\rho^2 \operatorname{div}(|x|^{\beta-2}x) + \beta\rho^2 \operatorname{div}(|x|^{\beta-\alpha-2}x)\\
&\quad = \beta\rho^2|x|^{\beta-\alpha-2}\left[2(1+|x|^\alpha)\nabla\log\rho\cdot x + (d+\beta-2)|x|^\alpha + d + \beta - \alpha - 2\right]\\
&\quad \leqslant \beta\rho^2|x|^{\beta-\alpha-2}\left[C(1+|x|^\alpha)|x| + (d+\beta-2)|x|^\alpha + d + \beta - \alpha - 2\right]
\end{aligned}$$

The proof is completed by choosing $c > 0$ small enough so that when $0 < |x| \leqslant c$

$$C(1+|x|^\alpha)|x| + (d+\beta-2)|x|^\alpha \leqslant \tfrac{1}{2}(\alpha+2-\beta-d). \qquad \square$$

Theorem 2.10 *Let $\alpha > d-2$. Assume γ satisfies (11), Ω has a $C^{2,\alpha}$ boundary, and $\rho \in C^{1,\sigma}(\overline{\Omega})$ is bounded above and below by positive constants . Then the elliptic Eq. (22) has a unique weak solution $u \in H^1_\gamma(\Omega)$. Furthermore, $u \in C(\overline{\Omega}) \cap C^{2,\sigma}_{loc}(\overline{\Omega}\backslash\Gamma)$ and satisfies for every $0 < \beta < \alpha + 2 - d$*

$$|u(x) - u(y)| \leqslant C(\beta)|x-y|^\beta \quad (x \in \overline{\Omega},\, y \in \Gamma). \tag{40}$$

Proof For $\varepsilon > 0$ set

$$\Omega_\varepsilon := \Omega\backslash \bigcup_{y\in\Gamma} \overline{B(y,\varepsilon)}$$

and let $u_\varepsilon \in C^{2,\sigma}(\overline{\Omega_\varepsilon})$ be the unique solution of the approximating problem

$$\begin{cases} -\operatorname{div}(\gamma\rho^2\nabla u_\varepsilon) = 0 & \text{in } \Omega_\varepsilon\\ u_\varepsilon = g(y) & \text{on } \partial B(y,\varepsilon) \text{ for all } y \in \Gamma\\ \dfrac{\partial u_\varepsilon}{\partial \nu} = 0 & \text{on } \partial\Omega. \end{cases} \tag{41}$$

It is a classical result that u_ε is the unique solution of the variational problem

$$\min\left\{\int_{\Omega_\varepsilon} \gamma\rho^2|\nabla u|^2\,dx \;:\; u \in H^1(\Omega_\varepsilon) \text{ and } \forall y\in\Gamma,\ u = g(y) \text{ on } \partial B(y,\varepsilon)\right\}. \tag{42}$$

In particular, it follows that

$$\sup_{\varepsilon>0} \int_{\Omega_\varepsilon} \gamma\rho^2 |\nabla u_\varepsilon|^2 \, dx < \infty. \tag{43}$$

By the maximum principle

$$\min_{\Gamma} g \leqslant u_\varepsilon \leqslant \max_{\Gamma} g. \tag{44}$$

Let $y \in \Gamma$. By Proposition 2.9, $w(x) := |x - y|^\beta$ satisfies

$$-\operatorname{div}(\gamma\rho^2 \nabla w) \geqslant \frac{\beta}{2}(\alpha + 2 - \beta - d)|x - y|^{-(\alpha+2-\beta)} > 0$$

for $0 < |x - y| \leqslant c$, where c depends on α, β, ρ, and d. Thus, another application of the maximum principle yields

$$u_\varepsilon(x) \leqslant g(y) + C|x - y|^\beta$$

for all $x \in \overline{\Omega_\varepsilon}$, where C is independent of $\varepsilon > 0$. The other direction is similar, yielding

$$|u_\varepsilon(x) - g(y)| \leqslant C|x - y|^\beta \quad \text{for all } x \in \overline{\Omega_\varepsilon}. \tag{45}$$

By the Schauder estimates [24], for each $\delta > 0$ there exists a constant $C > 0$, independent of ε, such that

$$\|u_\varepsilon\|_{C^{2,\sigma}(\overline{\Omega_\delta})} \leqslant C$$

for all $0 < \varepsilon < \delta$. Therefore, there exists a subsequence u_{ε_k} and $u \in C^{2,\sigma}_{loc}(\overline{\Omega}\backslash\Gamma)$ such that $u_{\varepsilon_k} \to u$ in $C^2_{loc}(\overline{\Omega}\backslash\Gamma)$. In particular, u solves (22) classically and satisfies (40), due to (45). Thus $u \in C(\overline{\Omega})$ and $u = g$ on Γ. Finally, it follows from (43) that $u \in H^1_\gamma(\Omega)$, and so u is a weak solution of (22), as per Definition 2.8. Uniqueness of weak solutions follows by a standard energy method argument. □

Lemma 2.11 *Let $\alpha > d - 2$. Assume γ satisfies (11), Ω has a $C^{2,\alpha}$ boundary, and $\rho \in C^{1,\sigma}(\overline{\Omega})$ is bounded above and below by positive constants . Then solution u of the variational problem (20) is the unique weak solution of the Euler–Lagrange equation (22).*

Proof Let $u \in H^1_\gamma(\Omega)$ be the unique weak solution of (22), and let $w \in H^1_\gamma(\Omega)$ with $\operatorname{Tr}[w] = \operatorname{Tr}[u]$. Then by the definition of weak solution

$$\int_\Omega \gamma\rho^2 \nabla u \cdot \nabla(u - w) \, dx = 0.$$

Therefore

$$\begin{aligned}\int_\Omega \gamma\rho^2|\nabla u|^2\,dx &= \int_\Omega \gamma\rho^2\nabla u\cdot\nabla w\,dx\\ &= \frac{1}{2}\int_\Omega \gamma\rho^2|\nabla u|^2\,dx + \frac{1}{2}\int_\Omega \gamma\rho^2|\nabla w|^2\,dx\\ &\quad - \frac{1}{2}\int_\Omega \gamma\rho^2|\nabla u-\nabla w|^2\,dx,\end{aligned}$$

and so

$$\int_\Omega \gamma\rho^2|\nabla u|^2\,dx = \int_\Omega \gamma\rho^2|\nabla w|^2\,dx - \int_\Omega \gamma\rho^2|\nabla u-\nabla w|^2\,dx.$$

It follows that u is the unique solution of the variational problem (20). □

3 Discrete to Continuum Convergence

Throughout this section we consider Ω, $\mathcal{X}_n$, and η which satisfy the assumptions of Sect. 1.2. Let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ be the empirical measure of the sample. Let $d_\infty(\mu,\mu_n)$ be the ∞-transportation distance between μ and μ_n, discussed in Appendix A.2.

We now state our main result. In order to compare discrete and continuum minimizers we use the TL^p topology introduced in [20]. We review the topology and its basic properties in Appendix A.3.

Theorem 3.1 *Let ε_n be a sequence of positive numbers converging to zero as $n\to\infty$ and such that $\varepsilon_n \gg d_\infty(\mu,\mu_n)$. Let $\zeta_n \in (1,\infty]$ be such that $\zeta_n \gg n\varepsilon_n^2$. Consider $\alpha > d-2$. Let u_n be a sequence of minimizers of the problem (14) for $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}$. Then almost surely (μ_n,u_n) converges in TL^2 to (μ,u) where u is the minimizer of (20).*

Our approach to proving the theorem is via establishing the Γ-convergence of the discrete constrained functionals to the continuum ones. The overall approach to consistency of learning algorithms follows the one developed in [20,22]. Ensuring that the discrete problem induces enough regularity for one to be able to show that the label values are preserved in the limit at points of Γ follows the general strategy of [41]. However the problems and proofs are rather different. We remark that one can also use the PDE-based approach of [10], but this would require a slightly more restrictive range on ε_n. Nevertheless the PDE-based approach gives superior regularity of solutions which we exploit in Sect. 4.

Proof Since, almost surely $\varepsilon_n \gg d_\infty(\mu,\mu_n) \gtrsim n^{-1/d}$ it follows that $\zeta_n\to\infty$ as $n\to\infty$. We note that by discrete comparison principle $\|u_n\|_{L^\infty(\mu_n)} \leqslant \max_\Gamma \|g\|$. By Lemma 3.5, the discrete energy $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}$ Γ-converges to $\theta_\eta\mathcal{E}$ and the sequence $\{(\mu_n,u_n)\}_{n=1,2,\dots}$ is precompact in TL^2. Therefore (μ_n,u_n) converges along a subsequence in TL^2 metric to $(\widetilde{\mu},u)$. Since μ_n converges to μ in Wasserstein metric, $\widetilde{\mu}=\mu$. The fact that u is the minimizer of (20) now follows directly from Γ-convergence of

Proposition 3.4 below. Consequently, the fact that the whole sequence u_n converges to u follows from the uniqueness of the minimizer of (20). □

Remark 3.2 While above we address only algebraically growing weights γ (see (11)) it is straightforward to modify the proofs to show that if γ grows faster than algebraically at labeled points (say $\gamma(x) = \exp(1/\operatorname{dist}(x, \Gamma))$ the conclusion of the theorem hold (in any dimension $d \geqslant 2$).

Remark 3.3 In this paper we assume that the data measure is supported on the set Ω of full dimension. There are no substantial obstacles in extending the results to the manifold setting where the data are sampled from a measure which is supported on a smooth submanifold of $\mathbb{R}^d$. One would only need to adjust the statements using manifold analogues of the weighted Dirichlet energy and the Laplacian. The convergence of graph Laplacian in the manifold setting has already been established in the standard setting [23]. In the manifold setting the dimension d in the results above should be replaced by the dimension of the data manifold.

Proposition 3.4 *Let ε_n be a sequence of positive numbers converging to zero as $n \to \infty$ and such that $\varepsilon_n \gg d_\infty(\mu, \mu_n)$. Let $\zeta_n \in (1, \infty]$ be such that $\zeta_n \gg n\varepsilon_n^2$ and $\zeta_n \gg \varepsilon_n^{2-d}$ if $d > 2$ and $\zeta_n \gg -\ln \varepsilon_n$ if $d = 2$. Let $\alpha > d - 2$. Then the constrained properly-weighted graph Dirichlet energy, defined on $TL^2(\Omega)$ by*

$$\mathcal{GE}^{con}_{n,\varepsilon_n,\zeta_n}(\widetilde{\mu}_n, u_n) = \begin{cases} \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n) & \text{if } \widetilde{\mu}_n = \mu_n \text{ and } u_n = g \text{ on } \Gamma \\ \infty & \text{else} \end{cases}$$

Γ-converges almost surely in TL^2 to the constrained continuum properly-weighted Dirichlet energy

$$\theta_\eta \mathcal{E}^{con}(\widetilde{\mu}, u) = \begin{cases} \theta_\eta \mathcal{E}(u) & \text{if } \widetilde{\mu} = \mu,\ u \in H^1_\gamma(\Omega) \text{ and } u = g \text{ on } \Gamma \\ \infty & \text{else}, \end{cases}$$

where the value of u on Γ is considered in the sense of the trace and

$$\theta_\eta = \frac{1}{d} \int_{\mathbb{R}^d} \eta(z)|z|^2 dz.$$

The proof of the Γ convergence of the unconstrained functionals follows from known results in a straightforward way. We state and prove it in a separate lemma below. The real difficulty is in proving that the constraints are preserved in the limit. Since the TL^2 topology alone is not sufficient to ensure this, we need to establish some control of oscillations near the labeled points. This relies on on several technical lemmas which are of some independent interest. We state them in the Sect. 3.2. The proof of Proposition 3.4 is presented in Sect. 3.3.

Lemma 3.5 *Assume $\alpha > 0$ and $\zeta_n \geqslant 1$. Under assumptions on $\mathcal{X}_n$ and ε_n of Proposition 3.4, the discrete energy $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}$ Γ-converges almost surely with respect to TL^2*

topology to the energy $\theta_\eta \mathcal{E}$, defined in (21) as $n \to \infty$ if $\zeta_n \to \infty$. Furthermore let $\{u_n\}_{n=1,2,\dots}$ be a sequence such that $\sup_n \mathcal{GE}_{n,\varepsilon_n,\zeta_n} < \infty$ and $\sup_n \|u_n\|_{L^\infty(\mu_n)} < \infty$. Then $\{(\mu_n, u_n) \,:\, n \in \mathbb{N}\}$ is precompact in TL^2.

Proof From results in the literature [20,21] it follows that for any fixed $\zeta > 0$ the discrete energies $\mathcal{GE}_{n,\varepsilon_n,\zeta}$ Γ-converge to $\theta_\eta \mathcal{E}(\cdot\,; \gamma_\zeta)$ as $n \to \infty$, under standard assumptions on ε_n. To show the liminf inequality for general ζ_n consider a sequence (μ_n, u_n) TL^2 converging to (μ, u). For any fixed k,

$$\liminf_{n\to\infty} \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n) \geqslant \theta_\eta \mathcal{E}(u; \zeta_k),$$

where $\mathcal{E}(u; \zeta_k)$ is given by (21) with γ replaced by γ_{ζ_k}, which implies the desired inequality by taking supremum over k. The limsup inequality follows by a simple diagonalization argument.

We recall from the literature (e.g [20] or Proposition 4.4 of [41]) that the precompactness of bounded sequences with bounded energies already holds for the weight $\gamma \equiv 1$. Thus the precompactness for $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}$ follows by comparison. □

3.1 The Negative Result

Proposition 3.6 *Let ε_n be a sequence of positive numbers converging to zero as $n \to \infty$ and such that $\varepsilon_n \gg d_\infty(\mu, \mu_n)$. Let $\zeta_n \geqslant 1$ be a sequence converging to infinity. Consider $\alpha \leqslant d - 2$. Then the constrained energy $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}$, defined in Proposition 3.4, Γ-converges almost surely in TL^2 metric to the unconstrained continuum energy $\theta_\eta \mathcal{E}$.*

Proof The liminf part of the Γ-convergence claim follows from the liminf claim of Lemma 3.5.

To show the limsup inequality, we first observe that by localizing near the points of Γ, and given that limsup inequality holds for the unconstrained functional, the problem can be reduced to considering $\Gamma = \{0\}$, $u \equiv 0$, and the construction a sequence of functions $u_n \in L^2(\mu_n)$ such that $u_n(0) = 1$, $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n) \to 0$ as $n \to \infty$ and $u_n \to 0$ in TL^2 as $n \to \infty$.

We now make some observation about the continuum functional. Namely when $\alpha \leqslant d - 2$ then the function $\varphi(x) = \ln\left(\ln\left(\frac{1}{|x|}\right)\right)$ belongs to $H^1_\gamma(B(0,1))$. Let $w_k = \max\{\min\{\frac{1}{k}\varphi(x), 1\}, 0\}$. Let $r_k > 0$ be such that $w_k = 1$ on $B(0, r_k)$. By mollifying we can obtain a smooth approximation v_k, $v_k = 1$ on $B(0, r_k/2)$ and $\|v_k\|_{H^1_\gamma(B(0,1))} \leqslant 2\|u_k\|_{H^1_\gamma(B(0,1))}$. Arguing as in Sect. 5 of [20], if one defines for each $k \in \mathbb{N}$, a sequence $u^k_n \in L^2(\mu_n)$ by $u^k_n(x_i) = v_k(x_i)$ for all $x_i \in X_n$ one has $u^k_n \to v_k$ in TL^2 and $\limsup_{n\to\infty} \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u^k_n) \geqslant \theta_\eta \mathcal{E}(v_k)$. Since $v_k \to 0$ in $H^1_\gamma(B(0,1))$ as $k \to \infty$, the conclusion follows by a diagonalization argument. □

Corollary 3.7 *Let ε_n be a sequence of positive numbers converging to zero as $n \to \infty$ and such that $\varepsilon_n \gg d_\infty(\mu, \mu_n)$. Let $\zeta_n \geqslant 1$ be a sequence converging to ∞ as $n \to \infty$. Let $\alpha \leqslant d - 2$. Let u_n be a sequence of minimizers of the problem (14)*

for $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}$. Let c_n be the average of u_n (with respect to measure μ_n). Then almost surely $(\mu_n, u_n - c_n)$ converges in TL^2 to $(\mu, 0)$; in other words the information about the labels is forgotten in the limit.

Proof Assume the claim is false. Then there exists $\delta > 0$ and a subsequence u_{n_j} such that foe all j, $d_{TL^2}((\mu_{n_j}, u_{n_j} - c_{n_j}), (\mu, 0)) > \delta$ for all j. By the maximum principle functions u_n are bounded by extremal values of g. Consequently, by Lemma 3.5, $u_{n_j} - c_{n_j}$ has a further convergent subsequence. Without a loss of generality we can assume that $u_{n_j} - c_{n_j}$ converges to some $v \in L^2(\mu)$. Then $\int v d\mu = \lim_{j\to\infty} \int u_{n_j} - c_{n_j} d\mu_{n_j} = 0$.

By the limsup part of Γ-convergence of Proposition 3.6 there exists a sequence $v_n \in L^2(\mu_n)$ such that $\mathcal{GE}^{con}_{n,\varepsilon_n,\zeta_n}(v_n) \to 0$ as $n \to \infty$. Since u_{n_j} are minimizers $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_{n_j} - c_{n_j}) \to 0$ as $n \to \infty$. We conclude by the liminf part of Γ-convergence that $\mathcal{E}(v) = 0$. Since $\int v d\mu = 0$ this implies that $v \equiv 0$, which contradicts the assumption about the sequence. □

We note that the analogue of the negative result in Corollary 3.7 for the standard graph Laplacian (corresponding to $\gamma \equiv 1$) was proved in [41][Theorem 2.1]. The following corollary then follows by the squeeze theorem for Γ-convergence.

Corollary 3.8 *Under the assumptions of Proposition 3.6 consider any sequence of graph based functionals $\mathcal{F}_n$ such that for $\mathcal{GE}_{n,\varepsilon_n,1} \leqslant \mathcal{F}_n \leqslant \mathcal{GE}_{n,\varepsilon_n,\zeta_n}$ (where we note that $\mathcal{GE}_{n,\varepsilon_n,1}$ is just a convenient way to write the standard graph Laplacian). Let u_n be the minimizers of (14) for $\mathcal{F}_n$ and let c_n be the average of u_n (with respect to measure μ_n). Then $(\mu_n, u_n - c_n)$ converges almost surely in TL^2 to $(\mu, 0)$.*

A particular consequence of this corollary is that the minimizers of the algorithm in [38] converge to a constant as $n \to \infty$.

3.2 Estimates for the Discrete to Continuum Convergence

Here we establish several results needed in the proofs of the main results above. We follow a similar strategy as [41]. Let us define the nonlocal continuum energy as

$$\mathcal{E}_{\varepsilon,\zeta}(u) = \frac{1}{\varepsilon^2} \iint \gamma_\zeta(x)\eta_\varepsilon(x-y)|u(x) - u(y)|^2 d\mu(x) d\mu(y). \tag{46}$$

It serves as an intermediary between the discrete graph based functionals and the continuum derivative-based functionals.

Lemma 3.9 (discrete to nonlocal control) *Consider Ω, μ, η, ζ, and x_i as in Theorem 3.1. Let $\widetilde{\eta}(|x|) = 1$ for $|x| \leqslant 1$ and $\widetilde{\eta}(|x|) = 0$ otherwise, and so $\widetilde{\eta} \leqslant \eta$. Let T_n be a sequence of transport maps satisfying the conclusions of Theorem A.3 and let $\widetilde{\varepsilon}_n = \varepsilon_n - 2\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}$. Define $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}(\cdot\,;\eta)$ by (13) and $\mathcal{E}_{\widetilde{\varepsilon}_n,\widetilde{\zeta}_n}(\cdot\,;\widetilde{\eta})$ be (46), where we explicitly denote the dependence of η. Let $\widetilde{\zeta}_n > 0$ be such that $\zeta_n \geqslant \widetilde{\zeta}_n$ and $\left(\frac{r_0}{2C}\right)^\alpha \ell_n^{-\alpha} \geqslant \widetilde{\zeta}_n$ where C is the constant from Theorem A.3 and ℓ_n is the*

transportation length scale from the same theorem. Then there exists $n_0 \in \mathbb{N}$ and a constant $\overline{C} > 0$ (independent of n and u_n) such that for all $n \geqslant n_0$

$$\mathcal{E}_{\widetilde{\varepsilon}_n,\widetilde{\zeta}_n}(u_n \circ T_n; \widetilde{\eta}) \leqslant \overline{C}\, \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n; \eta)$$

Proof If $\left|\frac{x-z}{\widetilde{\varepsilon}_n}\right| < 1$ then

$$|T_n(x) - T_n(z)| \leqslant 2\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)} + |x - z| \leqslant 2\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)} + \widetilde{\varepsilon}_n = \varepsilon_n.$$

So,

$$\left|\frac{x-z}{\widetilde{\varepsilon}_n}\right| < 1 \quad \text{implies} \quad \left|\frac{T_n(x) - T_n(z)}{\varepsilon_n}\right| \leqslant 1$$

and therefore

$$\left|\frac{x-z}{\widetilde{\varepsilon}_n}\right| < 1 \quad \text{implies} \quad \widetilde{\eta}\left(\frac{|x-z|}{\widetilde{\varepsilon}_n}\right) = 1 = \widetilde{\eta}\left(\frac{|T_n(x) - T_n(z)|}{\varepsilon_n}\right).$$

Hence,

$$\widetilde{\eta}\left(\frac{|x-z|}{\widetilde{\varepsilon}_n}\right) \leqslant \widetilde{\eta}\left(\frac{|T_n(x) - T_n(z)|}{\varepsilon_n}\right) \leqslant \eta\left(\frac{|T_n(x) - T_n(z)|}{\varepsilon_n}\right).$$

From the assumptions on $\widetilde{\zeta}_n$ and T_n follows that

$$2\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)} \leqslant r_0(\widetilde{\zeta}_n - 1)^{-\frac{1}{\alpha}} = r_{\widetilde{\zeta}_n}$$

where $r_{\widetilde{\zeta}_n}$ is the length scale such that $1 + \left(\frac{r_0}{\mathrm{dist}(x,\Gamma)}\right)^\alpha > \widetilde{\zeta}_n$ if $\mathrm{dist}(x,\Gamma) < r_{\widetilde{\zeta}_n}$. We claim that for a.e. $x \in \Omega$

$$\min\left\{1 + \left(\frac{r_0}{\mathrm{dist}(x,\Gamma)}\right)^\alpha, \widetilde{\zeta}_n\right\} \leqslant 2^\alpha \min\left\{1 + \left(\frac{r_0}{\mathrm{dist}(T_n(x),\Gamma)}\right)^\alpha, \widetilde{\zeta}_n\right\} \quad (47)$$

Namely if $d(x,\Gamma) \leqslant r_{\widetilde{\zeta}_n}$ then $d(T_n,\Gamma) \leqslant |T_n(x) - x| + r_{\widetilde{\zeta}_n} \leqslant 2r_{\widetilde{\zeta}_n}$ for a.e. such x. Thus

$$1 + \left(\frac{r_0}{\mathrm{dist}(T_n(x),\Gamma)}\right)^\alpha \geqslant 1 + \frac{1}{2^\alpha}\left(\frac{r_0}{r_{\widetilde{\zeta}_n}}\right)^\alpha \geqslant \frac{1}{2^\alpha}\widetilde{\zeta}_n$$

If $d(x,\Gamma) > r_{\widetilde{\zeta}_n}$ then $d(x,\Gamma) \geqslant \frac{1}{2}d(T_n(x),\Gamma)$ for a.e. such x. Thus $\left(\frac{r_0}{\mathrm{dist}(T_n(x),\Gamma)}\right)^\alpha \geqslant \frac{1}{2^\alpha}\left(\frac{r_0}{\mathrm{dist}(x,\Gamma)}\right)^\alpha$.

Using (47) we conclude

$$\begin{aligned}
&\mathcal{E}_{\widetilde{\varepsilon}_n,\widetilde{\zeta}_n}(u_n \circ T_n; \widetilde{\eta}) \\
&\quad = \frac{1}{\widetilde{\varepsilon}_n^2} \iint \min\left\{1 + \left(\frac{r_0}{\operatorname{dist}(x,\Gamma)}\right)^{\alpha}, \widetilde{\zeta}_n\right\} \widetilde{\eta}_{\widetilde{\varepsilon}_n}(x-y) \\
&\qquad |u_n(T_n(x)) - u_n(T_n(y))|^2 d\mu(x) d\mu(y) \\
&\quad \leqslant 2^{\alpha} \frac{\varepsilon_n^d}{\widetilde{\varepsilon}_n^{d+2}} \iint \min\left\{1 + \left(\frac{r_0}{\operatorname{dist}(T_n(x),\Gamma)}\right)^{\alpha}, \zeta_n\right\} \\
&\qquad \eta_{\varepsilon_n}(|T_n(x) - T_n(y)|)\, |u_n(T_n(x)) - u_n(T_n(z))|^2 \, d\mu(x) d\mu(y) \\
&\quad = 2^{\alpha} \frac{\varepsilon_n^{d+2}}{\widetilde{\varepsilon}_n^{d+2}} \mathcal{G}\mathcal{E}_{n,\varepsilon_n,\zeta_n}(u_n; \eta).
\end{aligned}$$

□

In the next lemma we show that boundedness of non-local energies implies regularity at scales greater than ε with weight $\gamma_{\widetilde{\zeta}}$. This allows us to relate non-local bounds to local bounds after mollification using a mollifier $J \in C_c^{\infty}(\mathbb{R}^d, [0,\infty))$, with $\int_{\mathbb{R}^d} J(x)\,dx = 1$, and $J_\varepsilon(x) = \frac{1}{\varepsilon^d} J(x/\varepsilon)$.

Lemma 3.10 (nonlocal to weak local control) *There exists a constant $C \geqslant 1$ and a radially symmetric mollifier J with $\operatorname{supp}(J) \subseteq \overline{B(0,1)}$ such that for all $\varepsilon > 0$, $u \in L^2(\Omega)$, and any $\Omega' \subset\subset \Omega$ (i.e. for every Ω' that is compactly contained in Ω) with $\operatorname{dist}(\Omega', \partial\Omega) > \varepsilon$ it holds that*

$$\mathcal{E}(u * J_\varepsilon; \gamma_{\widetilde{\zeta}}, \Omega') \leqslant C \mathcal{E}_{\varepsilon,\widetilde{\zeta}}(u; \Omega) \tag{48}$$

where for both functionals we explicitly denote the dependence of the domain.

Proof Let J be a radially symmetric mollifier whose support is contained in $\overline{B(0,1)}$. There exists $\beta > 0$ such that $J \leqslant \beta\eta(|\cdot|)$ and $|\nabla J| \leqslant \beta\eta(|\cdot|)$. Let $u_\varepsilon = J_\varepsilon * u$. For arbitrary $x \in \Omega$ with $\operatorname{dist}(x, \partial\Omega) > \varepsilon$ we have

$$\begin{aligned}
|\nabla u_\varepsilon(x)| &= \left| \int_\Omega \nabla J_\varepsilon\,(x-z)\, u(z)\, dz \right| \\
&= \left| \int_\Omega \nabla J_\varepsilon\,(x-z)\,(u(z) - u(x))\, dz - \int_{\mathbb{R}^d \setminus \Omega} \nabla J_\varepsilon\,(x-z)\, u(x)\, dz \right| \\
&\leqslant \frac{\beta}{\varepsilon^{d+1}} \int_\Omega \eta\left(\frac{|x-z|}{\varepsilon}\right) |u(z) - u(x)|\, dz \\
&\quad + \frac{1}{\varepsilon^{d+1}} \int_{\mathbb{R}^d \setminus \Omega} \left| \nabla J\left(\frac{x-z}{\varepsilon}\right) \right| |u(x)|\, dz.
\end{aligned}$$

where the second line follows from $\int_{\mathbb{R}^d} \nabla J(w)\, dw = 0$. For the second term we have

$$\frac{1}{\varepsilon^{d+1}} \int_{\mathbb{R}^d \setminus \Omega} \left| \nabla J\left(\frac{x-z}{\varepsilon}\right) \right| |u(x)|\, dz = 0$$

since for all $z \in \mathbb{R}^d \backslash \Omega$ and $x \in \Omega$ with $\operatorname{dist}(x, \partial\Omega) > \varepsilon$ it follows that $|x - z| > \varepsilon$ and thus $\nabla J\left(\frac{x-z}{\varepsilon}\right) = 0$. Therefore, for $\theta_\eta = \int_{\mathbb{R}^d} \eta(|w|)\, dw$,

$$\begin{aligned}
|\nabla u_\varepsilon(x)|^2 &\leqslant \beta^2 \left(\int_\Omega \frac{1}{\varepsilon} \eta_\varepsilon(|x - z|)\, |u(z) - u(x)|\, dz \right)^2 \\
&= \frac{\theta_\eta^2 \beta^2}{\varepsilon^2} \left(\int_\Omega \frac{\eta_\varepsilon(|x - z|)}{\theta_\eta} |u(z) - u(x)|\, dz \right)^2 \\
&\leqslant \theta_\eta \beta^2 \int_\Omega \eta_\varepsilon(|x - z|) \frac{|u(z) - u(x)|^2}{\varepsilon^2}\, dz
\end{aligned}$$

by Jensen's inequality (since $\frac{1}{\theta_\eta} \int_{\mathbb{R}^d} \eta_\varepsilon(|x - z|)\, dz = 1$). Hence,

$$\begin{aligned}
&\int_{\Omega'} |\nabla u_\varepsilon(x)|^2\, \gamma_{\widetilde{\zeta}}(x) \rho^2(x)\, dx \\
&\leqslant \theta_\eta \beta^2 \int_\Omega \int_\Omega \eta_\varepsilon(|x - z|) \left| \frac{u(z) - u(x)}{\varepsilon} \right|^2 \gamma_{\widetilde{\zeta}}(x) \rho^2(x)\, dz\, dx \\
&\leqslant \frac{\theta_\eta \beta^2 \sup_{x \in \Omega} \rho(x)}{\inf_{x \in \Omega} \rho(x)} \mathcal{E}_{\varepsilon, \widetilde{\zeta}}(u; \Omega)
\end{aligned}$$

which completes the proof. □

We now show that controlling the local energy with cut-off near the singularity is sufficient to be able to find a nearby (in H^1_γ) function which has a similarly bounded energy without a cut-off.

Lemma 3.11 (weak local to strong local control) *Consider $\widetilde{\zeta} > 1$ such that $r_{\widetilde{\zeta}}$ defined in (17) satisfies $r_{\widetilde{\zeta}} \leqslant \frac{1}{2} \min\{|x - y| : x, y \in \Gamma, x \neq y\}$. Let $\overline{r} = r_{\widetilde{\zeta}}$. Then there exists a constant $C > 0$ such that for every $u \in H^1(\Omega)$ there exists $v \in H^1_\gamma(\Omega)$ such that*

$$v|_{B(z, \overline{r}/2)} \equiv \frac{1}{|B(0, \overline{r}/2)|} \int_{B(z, \overline{r}/2)} u(x) dx \quad \text{for all } z \in \Gamma, \tag{49}$$

$$v = u \quad \text{on } \Omega \backslash \Gamma_{\overline{r}} \tag{50}$$

$$\mathcal{E}(v; \gamma) \leqslant C \mathcal{E}(u, \gamma_{\widetilde{\zeta}}). \tag{51}$$

Proof Using the finiteness of Γ, from Lemma A.4 via translations and a rescaling follows that there exists $c \geqslant 1$, and $v \in H^1(\Omega)$ satisfying (49) and (50) such that

$$\int_{\Gamma_{\overline{r}}} |\nabla v(x)|^2 dx \leqslant c \int_{\Gamma_{\overline{r}}} |\nabla u(x)|^2 dx.$$

Using that $\nabla v = 0$ on $\Gamma_{\frac{\bar r}{2}}$ and $1 + \left(\frac{r_0}{\bar r}\right)^\alpha = \widetilde{\zeta}$ we obtain

$$\begin{aligned}
\mathcal{E}(v;\gamma) &= \int_{\Gamma_{\bar r}} \gamma(x)|\nabla v(x)|^2\rho(x)dx + \int_{\Omega\setminus\Gamma_{\bar r}} \gamma(x)|\nabla u(x)|^2\rho(x)dx \\
&\leqslant \int_{\Gamma_{\bar r}\setminus\Gamma_{\frac{\bar r}{2}}} \left(1 + \left(\frac{2r_0}{\bar r}\right)^\alpha\right) |\nabla v(x)|^2\rho(x)dx + \int_{\Omega\setminus\Gamma_{\bar r}} \gamma_{\widetilde{\zeta}}|\nabla u(x)|^2\rho(x)dx \\
&\leqslant 2^\alpha \int_{\Gamma_{\bar r}\setminus\Gamma_{\frac{\bar r}{2}}} \widetilde{\zeta}\,|\nabla v(x)|^2\rho(x)dx + \int_{\Omega\setminus\Gamma_{\bar r}} \gamma_{\widetilde{\zeta}}|\nabla u(x)|^2\rho(x)dx \\
&\leqslant 2^\alpha c \int_{\Gamma_{\bar r}} \widetilde{\zeta}\,|\nabla u(x)|^2\rho(x)dx + \int_{\Omega\setminus\Gamma_{\bar r}} \gamma_{\widetilde{\zeta}}|\nabla u(x)|^2\rho(x)dx \\
&\leqslant 2^\alpha c \int_{\Omega} \gamma_{\widetilde{\zeta}}|\nabla u(x)|^2\rho(x)dx
\end{aligned}$$

□

Lemma 3.12 *There exists $C > 0$ such that for all $0 < \varepsilon < r \leqslant 1$, for all $u \in H^1(B(0,r))$ such that $u \leqslant 0$ on $B(0,\varepsilon)$*

$$\fint_{B(0,r)} u(x)dx \leqslant C\sqrt{\int_{B(0,r)\setminus B(0,\varepsilon)} |\nabla u|^2 dx}\begin{cases}\varepsilon^{\frac{2-d}{2}} & \text{if } d \geqslant 3\\ \sqrt{-\ln\varepsilon} & \text{if } d = 2.\end{cases} \tag{52}$$

Proof We only prove the claim for $r = 1$. The claim for general r follows by a simple change of variables $y = x/r$. Furthermore we note that we can assume that $u = 0$ on $B(0,\varepsilon)$, since for general u one can consider $\widetilde{u} = \max\{u, 0\}$ and note that

$$\begin{aligned}
\int_{B(0,1)} u dx \leqslant \int_{B(0,1)} \widetilde{u} dx \quad \text{and} \quad &\sqrt{\int_{B(0,1)\setminus B(0,\varepsilon)} |\nabla\widetilde{u}|^2 dx} \\
&\leqslant \sqrt{\int_{B(0,1)\setminus B(0,\varepsilon)} |\nabla u|^2 dx}.
\end{aligned}$$

Let $v(x) = \frac{1}{2d}|x|^2$ and

$$\Phi(x) = \begin{cases} -\frac{1}{2}\log|x|, & \text{if } d = 2\\ \frac{1}{d(d-2)|x|^{d-2}}, & \text{if } d \geqslant 3.\end{cases}$$

Then $\Delta(v + \Phi) = 1$ for $x \neq 0$ and $\frac{\partial v}{\partial \nu} + \frac{\partial \Phi}{\partial \nu} = 0$ on $\partial B(0,1)$. Since $u = 0$ on $B(0,\varepsilon)$ we have

$$\begin{aligned}
\int_{B(0,1)} u\,dx &= \int_{B(0,1)\setminus B(0,\varepsilon)} u\Delta(v+\Phi)\,dx \\
&= -\int_{B(0,1)\setminus B(0,\varepsilon)} \nabla u\cdot(\nabla v + \nabla\Phi)\,dx
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{d} \int_{B(0,1)\setminus B(0,\varepsilon)} \nabla u \cdot x \left(\frac{1}{|x|^d} - 1 \right) dx \\
&\leqslant \frac{2}{d} \int_{B(0,1)\setminus B(0,\varepsilon)} |\nabla u| |x|^{1-d} \, dx \\
&\leqslant \frac{2}{d} \sqrt{\int_{B(0,1)\setminus B(0,\varepsilon)} |\nabla u|^2 \, dx} \sqrt{\int_{B(0,1)\setminus B(0,\varepsilon)} |x|^{2-2d} \, dx}.
\end{aligned}
$$

The proof is completed by integrating the last term on the line above. □

3.3 Proof of Proposition 3.4

Proof To show the lim sup inequality recall that $\mathcal{S}$, the set of smooth functions which are constant in some neighborhood of Γ, is dense in $H^1_\gamma(\Omega)$, by Corollary 2.5. The fact that for every $f \in \mathcal{S}$, $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}(f) \to \theta_\eta \mathcal{E}(f)$ follows by a standard argument, which was for example presented for total variation in Sect. 5 of [20]. The existence of a recovery sequence for arbitrary $f \in H^1_\gamma(\Omega)$ follows by a density argument.

To show the lim inf inequality consider a sequence (μ_n, u_n) converging in TL^2 to (μ, u). We can assume without a loss of generality that $u_n|_\Gamma = g$ and that $\liminf_{n\to\infty} \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n)$ is finite. Since $\varepsilon_n \gg d_\infty(\mu, \mu_n) \gtrsim n^{-1/d}$ it follows that $\zeta_n \to \infty$ as $n \to \infty$. By Lemma 3.5, discrete energy $\mathcal{GE}_{n,\varepsilon_n,\zeta_n}$ Γ-converges to $\mathcal{E}$. Thus $u \in H^1_\gamma(\Omega)$ and

$$
\liminf_{n\to\infty} \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n) \geqslant \theta_\eta \mathcal{E}(u).
$$

What remains to be shown is that $u|_\Gamma = g$. The fact that $u|_\Gamma$ is a well defined object follows from Lemma 2.1. Let us assume that $\liminf_{n\to\infty} \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n) = \lim_{n\to\infty} \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n)$. For the general case one needs to consider a subsequence, which we omit for notational simplicity. We have that $E_{max} = \sup_n \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n) < \infty$.

We first show that near points $z \in \Gamma$, the values of u_n remain, on average, close to $g(z)$. More precisely

$$
E_{max} \geqslant \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n) \geqslant \frac{1}{2n^2\varepsilon_n^2} \sum_{x\in\mathcal{X}_n} \zeta_n \eta_{\varepsilon_n}(x - z) |u_n(x) - g(z)|^2.
$$

and thus

$$
\frac{1}{n} \sum_{x\in\mathcal{X}_n} \eta_{\varepsilon_n}(x - z) |u_n(x) - g(z)|^2 \leqslant 2E_{max} \frac{n\varepsilon_n^2}{\zeta_n}.
$$

Since $\eta \geqslant 1$ on $B(0,1)$

$$\frac{1}{n}\sum_{x\in\mathcal{X}_n,|x-z|<\varepsilon_n}|u_n(x)-g(z)|^2 \leqslant 2E_{max}\,\frac{n\varepsilon_n^{d+2}}{\zeta_n}.$$

Let T_n be a sequence of transport maps satisfying the conclusions of Theorem A.3 and let $\widetilde{\varepsilon}_n = \varepsilon_n - 2\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}$.

Then for a.e. $x \in B(z,\widetilde{\varepsilon}_n)$, $T_n(x) \in B(z,\varepsilon_n)$ and thus

$$\int_{B(z,\widetilde{\varepsilon}_n)}|u_n(T_n(x))-g(z)|^2\rho(x)dx \leqslant 2E_{max}\,\frac{n\varepsilon_n^{d+2}}{\zeta_n}.$$

Therefore

$$⨍_{B(z,\widetilde{\varepsilon}_n)}|u_n(T_n(x))-g(z)|^2\rho(x)dx \lesssim \frac{n\varepsilon_n^2}{\zeta_n} \ll 1 \tag{53}$$

by the assumption on ζ_n. Consequently for all $y \in B\left(z,\frac{\widetilde{\varepsilon}_n}{2}\right)$

$$⨍_{B(y,\widetilde{\varepsilon}_n/2)}|u_n(T_n(x))-g(z)|^2dx \lesssim \frac{n\varepsilon_n^2}{\zeta_n} \ll 1. \tag{54}$$

By Lemma 3.9 we know that, for $\widetilde{\zeta}_n = \min\left\{\zeta_n, \left(\frac{r_0}{2C}\right)^\alpha\left(\frac{n}{\ln n}\right)^{\alpha/d}\right\}$ where C is the constant from Theorem A.3

$$\mathcal{E}_{\widetilde{\varepsilon}_n,\widetilde{\zeta}_n}(u_n\circ T_n;\widetilde{\eta}) \lesssim \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n;\eta) \tag{55}$$

We note that since $\alpha > d-2$, and $\varepsilon_n \gg \left(\frac{\ln n}{n}\right)^{1/d}$, $\widetilde{\zeta}_n \gg \varepsilon_n^{2-d}$ if $d > 2$ and $\widetilde{\zeta}_n \gg -\ln\varepsilon_n$ Let $\widehat{\varepsilon}_n = \frac{\widetilde{\varepsilon}_n}{2}$. By definition of $\mathcal{E}_{\widehat{\varepsilon}_n,\widetilde{\zeta}_n}$

$$\mathcal{E}_{\widehat{\varepsilon}_n,\widetilde{\zeta}_n}(u_n\circ T_n;\widetilde{\eta}) \lesssim \mathcal{E}_{\widetilde{\varepsilon}_n,\widetilde{\zeta}_n}(u_n\circ T_n;\widetilde{\eta}) \tag{56}$$

Let J be a mollifier used in the proof of Lemma 3.10 and let $\widetilde{u}_n = (u_n\circ T_n) * J_{\widehat{\varepsilon}_n}$. From (54) follows that for all $y \in B(z,\widehat{\varepsilon}_n)$

$$|\widetilde{u}_n(y)-g(z)| \lesssim \frac{n\varepsilon_n^2}{\zeta_n} \tag{57}$$

for all $z \in \Gamma$. Combining the estimate of the lemma with (55) yields

$$\mathcal{E}(\widetilde{u}_n;\gamma_{\widetilde{\zeta}_n},\Omega_n') \lesssim \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n;\eta)$$

where $\Omega_n' = \{y\in\Omega : d(y,\partial\Omega) > \varepsilon_n\}$. Finally by Lemma 3.11 there exist $v_n \in H^1_\gamma(\Omega_n')$ such that for all $z\in\Gamma$, $v_n(z) = ⨍_{B(z,\overline{r}/2)}\widetilde{u}_n(y)dy$ where $\overline{r} = r_{\widetilde{\zeta}_n}$ and

$$\mathcal{E}(v_n, \Omega_n') \lesssim \mathcal{GE}_{n,\varepsilon_n,\zeta_n}(u_n; \eta). \tag{58}$$

From (57) follows that for some C independent of n, for all $z \in \Gamma$, and all $y \in B(z, \widehat{\varepsilon}_n)$, using that $\int_{B(0,\bar{r}/2)} |\nabla \widetilde{u}_n|^2 dx \leqslant \frac{1}{\zeta_n}\mathcal{E}(\widetilde{u}_n; \gamma_{\zeta_n}, \Omega_n')$, $u(y) - g(z) - C\frac{n\varepsilon_n^2}{\zeta_n} \leqslant 0$. Thus by Lemma 3.12

$$\fint_{B(0,\bar{r}/2)} \widetilde{u}_n(x)dx \leqslant g(z) + C\frac{n\varepsilon_n^2}{\zeta_n} + C_1\widetilde{\zeta}_n^{-1/2}\begin{cases}\varepsilon^{\frac{2-d}{2}} & \text{if } d \geqslant 3\\ \sqrt{-\ln\varepsilon} & \text{if } d = 2.\end{cases}$$

for some C_1 independent of n. By the assumptions on ζ_n and definition of $\widetilde{\zeta}_n$ the right hand side converges to zero an $n \to \infty$. Analogous lower bound is obtained following the same argument. Therefore $v_n(z) - g(z) = \fint_{B(0,\bar{r}/2)} \widetilde{u}_n(x)dx - g(z)$ converges to zero as $n \to \infty$.

We note that by construction $d_{TL^2}((\mu_n, u_n), (\mu, \widetilde{u}_n)) \to 0$ and thus $\widetilde{u}_n \to u$ in $L^2(\Omega)$. By construction $\|v_n - \widetilde{u}_n\|_{L^2(\Omega_n')} \lesssim \|\widetilde{u}_n\|_{L^2(\Gamma_{\bar{r}_n})}$, where $\bar{r}_n = r_{\widetilde{\zeta}_n}$ and for any $s > 0$

$$\Gamma_s = \{x \in \Omega \ : \ \operatorname{dist}(x, \Gamma) < s\}. \tag{59}$$

Since $(\mu_n, u_n) \xrightarrow{TL^2} (\mu, u)$ it follows that

$$\int_{T_n^{-1}(\Gamma_{\bar{r}_n})} |u_n(T_n(x)) - u(x)|^2 d\mu(x) \to 0 \qquad \text{as } n \to \infty.$$

Since $u \in L^2(\mu)$, $\lim_{\delta\to 0} \sup\left\{\int_A u^2(x)dx \ : \ \mu(A) < \delta\right\} = 0$. Therefore

$$\|\widetilde{u}_n\|_{L^2(\Gamma_{\bar{r}_n})} \leqslant 2\int_{T_n^{-1}(\Gamma_{\bar{r}_n})} |u_n(T_n(x)) - u(x)|^2 + u^2(x)d\mu(x) \to 0 \qquad \text{as } n \to \infty.$$

Thus $\|v_n - \widetilde{u}_n\|_{L^2(\Omega_n')} \to 0$ and $n \to \infty$ and consequently $v_n \to u$ in $L^2(\mu)$. From (58) follows that v_n is a bounded sequence in $H^1_\gamma(K)$ for any compact subset $K \subset\subset \Omega$. Combining this with the fact that $v_n \to u$ in $L^2(\mu)$ implies, via estimate (34) of the Trace Theorem (Theorem 2.2), that $v_n(z) \to \operatorname{Tr} u(z)$ as $n \to \infty$ for all $z \in \Gamma$. Since $v_n|_\Gamma \to g$ as $n \to \infty$ we conclude that $\operatorname{Tr} u(z) = g(z)$ for all $z \in \Gamma$. □

4 Regularity of Minimizers of the Graph Properly-Weighted Laplacian

4.1 Hölder Estimate Near Labeled Points

Our main result in this section is a type of Hölder estimate near the labeled points, which shows that solutions of the graph-based learning problem (15) attain their boundary values on Γ continuously, with high probability. The proof is a graph-based version of the barrier argument from Theorem 2.10 that established continuity at labels in the

continuum PDE, given in Eq. (40). Barrier arguments for proving Hölder regularity of solutions of PDEs are standard techniques for first order equations, such as Hamilton–Jacobi equations [4]. Normally, barrier arguments do not work for second order elliptic equations (since fundamental solutions are unbounded), though there are a handful of exceptions, such as the p-Laplace equation for $p > d$ [10], level set equations for affine curvature motion [12], and our continuum equation (22).

Our proof uses the barrier $v(x) = C|x - y|^\beta$, which is a supersolution of the continuum PDE (22) for $0 < \beta < \alpha + 2 - d$, due to Proposition 2.9. For $\beta < 2$, the barrier has a singularity at $x = y$, which has to be treated carefully in the translation to the graph setting. We show in Lemma 4.6 that v is a supersolution on the graph with high probability away from a small ball $B(y, C\varepsilon)$. Due to the singularity in the barrier, we cannot prove the supersolution property within the ball $B(y, C\varepsilon)$. To fill in the gap within this ball, we require a local regularity result, given in Lemma 4.8, that relies on the variational structure of the problem. At a high level, the proof is similar to the proof of Hölder regularity of solutions to the graph-based game theoretic p-Laplace equation, given in [10], though many of the ingredients are different. In particular, in [10] there is no variational interpretation of the problem, and the local argument utilizes another barrier construction.

We now proceed to present the main results in this section. Throughout we always assume $n\varepsilon^d \geqslant 1$. Our main result is the following Hölder-type estimate.

Theorem 4.1 (Hölder estimate) *Let $\alpha > d - 2$, and assume γ satisfies (11) and $\rho \in C^{1,\sigma}(\overline{\Omega})$ is bounded above and below by positive constants. Let $0 < \varepsilon \leqslant 1$, $\zeta \geqslant 1 + \varepsilon^{-\alpha}$, $0 < \beta < \alpha + 2 - d$, and let $u \in L^2(\mathcal{X}_n)$ be the solution of (15). Then for each $z \in \Gamma$ the event that*

$$|u(x_i) - u(z)| \leqslant C|x_i - z|^\beta + Cn^{1/2}\varepsilon^{1+\alpha/2} \tag{60}$$

holds for all $x_i \in \mathcal{X}_n$ occurs with probability at least $1 - C\exp\left(-cn\varepsilon^{d+4} + \log(n)\right)$.

The proof of Theorem 4.1, given at the end of the section, relies on some preliminary results that we establish after a few remarks.

Remark 4.2 For the result in Theorem 4.1 to be useful, we must choose $\varepsilon_n \to 0$ so that $n\varepsilon_n^{d+4} \gg \log(n)$ and $n\varepsilon_n^{\alpha+2} \ll 1$. Therefore, we must have $\alpha > d + 2$ and

$$\left(\frac{\log(n)}{n}\right)^{1/(d+4)} \ll \varepsilon_n \ll \left(\frac{1}{n}\right)^{1/(\alpha+2)}. \tag{61}$$

Remark 4.3 If we replace γ_ζ with $\gamma_{\zeta, C\varepsilon}$, as defined in Remark 1.3, then we can improve Theorem 4.1 to read

$$|u(x_i) - u(z)| \leqslant C|x_i - z|^\beta + C\zeta^{-1/2}n^{1/2}\varepsilon, \tag{62}$$

under the same assumptions and with the same probability, except we also require $\zeta \geqslant 1 + C\varepsilon^{-\alpha}$. In this model, the restrictive upper bound in (61) is not required.

We now turn to the proof of Theorem 4.1. We first recall a useful lemma from [10].

Lemma 4.4 (Remark 7 from [10]) *Let $Y_1, Y_2, Y_3, \ldots, Y_n$ be a sequence of* i.i.d *random variables on $\mathbb{R}^d$ with Lebesgue density $\rho : \mathbb{R}^d \to \mathbb{R}$, let $\psi : \mathbb{R}^d \to \mathbb{R}$ be bounded and Borel measurable with compact support in a ball $B(x, h)$ for some $h > 0$, and define*

$$Y = \sum_{i=1}^{n} \psi(Y_i).$$

Then for any $0 \leqslant \lambda \leqslant 1$

$$\mathbb{P}\left(|Y - \mathbb{E}(Y)| \geqslant C\|\psi\|_{L^\infty(B(x,h))} nh^d \lambda\right) \leqslant 2\exp(-cnh^d\lambda^2), \tag{63}$$

for all $0 < \lambda \leqslant 1$, where $C, c > 0$ are constants depending only on $\|\rho\|_{L^\infty(B(x,h))}$ and d.

We can use Lemma 4.4 to prove pointwise consistency for our properly-weighted graph Laplacian. It extends, in a refined form, the results of [10][Theorem 5]. It is related to well known results on the pointwise consistency of the graph Laplacian [40].

For simplicity we set

$$\Delta_\rho \varphi = \rho^{-1} \operatorname{div}\left(\gamma \rho^2 \nabla \varphi\right). \tag{64}$$

Theorem 4.5 *For $\delta > 0$, let $D_{n,\varepsilon,\delta}$ be the event that*

$$\begin{aligned}\left|\mathcal{GL}_{n,\varepsilon,\infty}\varphi(x_i) - \tfrac{1}{2}\sigma_\eta \Delta_\rho \varphi(x_i)\right| \leqslant C((\varepsilon\beta_1 + \varepsilon^2\beta_2)M^{-(\alpha+2)} \\ + \varepsilon\beta_2 M^{-(\alpha+1)} + \delta\beta_3 M^{-\alpha})\end{aligned} \tag{65}$$

holds for all x_i with $2\varepsilon < \operatorname{dist}(x_i, \Gamma) \leqslant R/4$ and all $\varphi \in C^3(B(x_i, 2\varepsilon))$, where $\beta_k = \|\varphi\|_{C^k(B(x_i,2\varepsilon))}$ and $M = \operatorname{dist}(x_i, \Gamma) - 2\varepsilon$. Then for $\varepsilon \leqslant \delta \leqslant 1$ we have

$$\mathbb{P}(D_{n,\varepsilon,\delta}) \geqslant 1 - C\exp\left(-c\delta^2 n\varepsilon^{d+2} + \log(n)\right) \tag{66}$$

Proof Let us write $\mathcal{GL}$ in place of $\mathcal{GL}_{n,\varepsilon,\infty}$ for simplicity. We also define

$$w(x, y) = \frac{1}{2n\varepsilon^2}(\gamma(x) + \gamma(y))\eta_\varepsilon(x - y).$$

Then

$$\mathcal{GL}u(x) = \sum_{y \in \mathcal{X}_n} w(x, y)(u(y) - u(x)).$$

By conditioning on the location of $x \in X_n$, we can assume without loss of generality that $x \in \Omega$ is a fixed (non-random) point, $B(x, 2\varepsilon) \subset \Omega$ and $\mathrm{dist}(x, \Gamma) > 2\varepsilon$. Let $\varphi \in C^3(B(x, 2\varepsilon))$, $p = D\varphi(x)$ and $A = D^2\varphi(x)$. Note that

$$\mathcal{G}\mathcal{L}\varphi(x) = \sum_{i=1}^{d} p_i \sum_{y \in X_n} w(x, y)(y_i - x_i) + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij} \sum_{y \in X_n} w(x, y)(y_i - x_i)(y_j - x_j) + O\left(\varepsilon^3 \beta_3 \deg(x)\right), \tag{67}$$

where $\deg(x)$ is the degree given by

$$\deg(x) = \sum_{y \in \mathcal{X}_n} w(x, y).$$

Since $\mathrm{dist}(y, \Gamma) \geqslant \mathrm{dist}(x, \Gamma) - 2\varepsilon$ we have

$$\begin{aligned} w(x, y) &\leqslant \frac{C}{2n\varepsilon^{d+2}} \left(\frac{1}{\mathrm{dist}(x, \Gamma)^\alpha} + \frac{1}{\mathrm{dist}(y, \Gamma)^\alpha} \right) \\ &\leqslant \frac{C}{n\varepsilon^{d+2}(\mathrm{dist}(x, \Gamma) - 2\varepsilon)^\alpha}. \end{aligned}$$

By Lemma 4.4

$$\left| \deg(x) - n \int_{B(x,2\varepsilon)} w(x, y)\rho(y)\, dy \right| \leqslant C\varepsilon^{-2}(\mathrm{dist}(x, \Gamma) - 2\varepsilon)^{-\alpha},$$

holds with probability at least $1 - 2\exp\left(-cn\varepsilon^d\right)$. This implies

$$\deg(x) \leqslant C\varepsilon^{-2}(\mathrm{dist}(x, \Gamma) - 2\varepsilon)^{-\alpha}.$$

By another application of Lemma 4.4, both

$$\begin{aligned} &\left| \sum_{y \in X_n} w(x, y)(y_i - x_i) - n \int_{B(x,2\varepsilon)} w(x, y)(y_i - x_i)\rho(y)\, dy \right| \\ &\quad \geqslant C\delta(\mathrm{dist}(x, \Gamma) - 2\varepsilon)^{-\alpha}, \end{aligned}$$

and

$$\begin{aligned} &\left| \sum_{y \in X_n} w(x, y)(y_i - x_i)(y_j - x_j) - n \int_{B(x,2\varepsilon)} w(x, y)(y_i - x_i)(y_j - x_j)\rho(y)\, dy \right| \\ &\quad \geqslant C\delta\varepsilon(\mathrm{dist}(x, \Gamma) - 2\varepsilon)^{-\alpha}, \end{aligned}$$

occur with probability at most $2\exp\left(-c\delta^2 n\varepsilon^{d+2}\right)$ provided $0 < \delta\varepsilon \leqslant 1$. Thus, if $\varepsilon \leqslant \delta \leqslant \varepsilon^{-1}$ we have

$$\begin{aligned}\mathcal{G}\mathcal{L}\varphi(x) = \frac{1}{2}\int_{B(0,2)} (\gamma(x)+\gamma(x+z\varepsilon))\rho(x+z\varepsilon)\boldsymbol{\eta}(|z|)\left(\frac{1}{\varepsilon}p\cdot z+\frac{1}{2}z\cdot Az\right)dz \\ + O\left(\delta\beta_3(\operatorname{dist}(x,\Gamma)-2\varepsilon)^{-\alpha}\right)\end{aligned} \tag{68}$$

holds for all $\varphi \in C^3(\mathbb{R}^d)$ with probability at least $1 - C\exp\left(-c\delta^2 n\varepsilon^{d+2}\right)$. Note that

$$\begin{aligned}\frac{1}{2}(\gamma(x)+\gamma(x+z\varepsilon))\rho(x+z\varepsilon) = \gamma(x)\rho(x)+\gamma(x)\nabla\rho(x)\cdot z\varepsilon+\frac{1}{2}\rho(x)\nabla\gamma(x)\cdot z\varepsilon \\ + O(\varepsilon^2\|\gamma\|_{C^2(B(x,2\varepsilon))}+\varepsilon^3\|\gamma\|_{C^1(B(x,2\varepsilon))}).\end{aligned}$$

We now have

$$\begin{aligned}&\int_{B(0,2)}\gamma(x)\rho(x)\boldsymbol{\eta}(|z|)\left(\frac{1}{\varepsilon}p\cdot z+\frac{1}{2}z\cdot Az\right)dz\\ &\quad=\frac{1}{2}\gamma(x)\rho(x)\sum_{i,j=1}^{d}a_{ij}\int_{B(0,2)}\boldsymbol{\eta}(|z|)z_iz_j\,dz\\ &\quad=\frac{1}{2}\gamma(x)\rho(x)\sum_{i=1}^{d}a_{ii}\int_{B(0,2)}\boldsymbol{\eta}(|z|)z_i^2\,dz\\ &\quad=\frac{\sigma_\eta}{2}\gamma(x)\rho(x)\operatorname{Trace}(A),\\ &\int_{B(0,2)}\gamma(x)(\nabla\rho(x)\cdot z\varepsilon)\boldsymbol{\eta}(|z|)\left(\frac{1}{\varepsilon}p\cdot z+\frac{1}{2}z\cdot Az\right)dz\\ &\quad=\gamma(x)\nabla\rho(x)\cdot\int_{B(0,2)}\boldsymbol{\eta}(|z|)(p\cdot z)z\,dz+O(\varepsilon\beta_2\operatorname{dist}(x,\Gamma)^{-\alpha})\\ &\quad=\gamma(x)\nabla\rho(x)\cdot\sum_{i=1}^{d}p_i\int_{B(0,2)}\boldsymbol{\eta}(|z|)z_iz\,dz+O(\varepsilon\beta_2\operatorname{dist}(x,\Gamma)^{-\alpha})\\ &\quad=\sigma_\eta\gamma(x)\nabla\rho(x)\cdot p+O(\varepsilon\beta_2\operatorname{dist}(x,\Gamma)^{-\alpha}),\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{2}\int_{B(0,2)}\rho(x)(\nabla\gamma(x)\cdot z\varepsilon)\boldsymbol{\eta}(|z|)\left(\frac{1}{\varepsilon}p\cdot z+\frac{1}{2}z\cdot Az\right)dz\\ &\quad=\frac{\sigma_\eta}{2}\rho(x)\nabla\gamma(x)\cdot p+O(\varepsilon\beta_2\|\gamma\|_{C^1(B(x,2\varepsilon))}).\end{aligned}$$

Assembling these together with (68) we have that

$$\mathcal{GL}\varphi(x) = \frac{\sigma_\eta}{2\rho(x)} \operatorname{div}\left(\gamma\rho^2\nabla\varphi\right) + O\Big(\varepsilon\beta_1\|\gamma\|_{C^2(B(x,2\varepsilon))} + \varepsilon\beta_2\|\gamma\|_{C^1(B(x,2\varepsilon))} + \varepsilon^2\beta_2\|\gamma\|_{C^2(B(x,2\varepsilon))} + \delta\beta_3(\operatorname{dist}(x,\Gamma) - 2\varepsilon)^{-\alpha}\Big)$$

holds for all $\varphi \in C^3(B(x, 2\varepsilon))$ with probability at least $1 - C\exp\left(-c\delta^2 n\varepsilon^{d+2}\right)$. The proof is completed by computing

$$\|\gamma\|_{C^k(B(x,2\varepsilon))} \leqslant C(\operatorname{dist}(x,\Gamma) - 2\varepsilon)^{-\alpha-k},$$

and applying a union bound over $x_1, \ldots, x_n$. □

We now establish that the function $|x|^\beta$ for $0 < \beta < \alpha + 2 - d$ serves as a barrier (e.g., is a supersolution) on the graph with high probability.

Lemma 4.6 (Barrier lemma) *Let $\alpha > d-2$ and fix any $0 < \beta < \alpha+2-d$. For $y \in \Gamma$ define $\varphi(x) = |x - y|^\beta$. Then the event that*

$$\mathcal{GL}_{n,\varepsilon,\infty}\varphi(x_i) \leqslant -c|x_i - y|^{-(\alpha+2-\beta)} \tag{69}$$

for all x_i with $C\varepsilon < |x_i - y| \leqslant c$ occurs with probability at least $1 - C\exp\left(-cn\varepsilon^{d+4} + \log(n)\right)$.

Proof Let us write $\mathcal{GL}$ in place of $\mathcal{GL}_{n,\varepsilon,\infty}$ for simplicity. We use Theorem 4.5 and Proposition 2.9. Note in Theorem 4.5 that if we restrict $3 \leqslant \varepsilon|x_i - y| \leqslant R/4$ then

$$M = \operatorname{dist}(x_i,\Gamma) - 2\varepsilon = |x_i - y| - 2\varepsilon \geqslant \frac{1}{3}|x_i - y|.$$

Also, for $\beta_k = \|\varphi\|_{C^k(B(x_i,2\varepsilon))}$ we compute

$$\beta_k \leqslant C\beta|x_i - y|^{\beta-k}.$$

Hence, setting $\delta = \varepsilon$ in Theorem 4.5 we obtain that

$$\left|\mathcal{GL}\varphi(x_i) - \tfrac{1}{2}\sigma_\eta\Delta_\rho\varphi(x_i)\right| \leqslant C\varepsilon|x_i - y|^{\beta-\alpha-3}(1 + \varepsilon|x_i - y|^{-1}) \tag{70}$$

holds for all x_i with $3\varepsilon \leqslant |x_i - y| \leqslant r$ with probability at least

$$1 - C\exp\left(-cn\varepsilon^{d+4} + \log(n)\right).$$

For the rest of the proof we restrict to the event that (70) holds.

Note that since $\beta - \alpha - 3 < 0$ and $|x_i - y| \geqslant 3\varepsilon$, it follows from (70) that

$$\mathcal{GL}\varphi(x_i) \leqslant \tfrac{1}{2}\sigma_\eta\Delta_\rho\varphi(x_i) + C\varepsilon|x_i - y|^{\beta-\alpha-3}. \tag{71}$$

Combining this with Proposition 2.9 we have

$$\begin{aligned}\mathcal{GL}\varphi(x_i) &\leqslant -c|x_i - y|^{\beta-\alpha-2} + C\varepsilon|x_i - y|^{\beta-\alpha-3} \\ &= -c|x_i - y|^{\beta-\alpha-2}(1 - C\varepsilon|x_i - y|^{-1}),\end{aligned}$$

provided $3\varepsilon \leqslant |x_i - y| \leqslant c$. The proof is completed by restricting $|x_i - y| \geqslant 2C\varepsilon$. □

The barrier lemma (Lemma 4.6) establishes the barrier property away from the local neighborhood $B(x, C\varepsilon)$. The singularity in the barrier (for $\alpha < d$) and the singularity in γ prevent us from pushing the barrier lemma inside this local neighborhood. Hence, the barrier can only be used to establish the following macroscopic continuity result.

Proposition 4.7 (Macroscopic Hölder estimate) *Let $u \in L^2(\mathcal{X}_n)$ be the solution of (15), let $\alpha > d - 2$, and fix any $0 < \beta < \alpha + 2 - d$. For each $y \in \Gamma$ the event that*

$$u(x_i) - u(y) \leqslant C|x_i - y|^{\beta} + \sup_{x \in \mathcal{X}_n \cap B(y, \delta_{\varepsilon,\zeta})} (u(x) - u(y)) \tag{72}$$

holds for all $x_i \in \mathcal{X}_n$ occurs with probability at least $1 - C\exp\left(-cn\varepsilon^{d+4} + \log(n)\right)$, where $\delta_{\varepsilon,\zeta} = \max\{C\varepsilon, r_\zeta + 2\varepsilon\}$.

Proof We note the graph is connected with probability at least $1 - C\exp(-cn\varepsilon^d + \log(n))$. The proof uses the barrier function

$$\varphi(x) = \sup_{B(y,\delta_{\varepsilon,\zeta})} u + K|x - y|^{\beta} \tag{73}$$

constructed in Lemma 4.6 for a sufficiently large K, and the maximum principle on a connected graph. By Lemma 4.6 we have

$$\mathcal{GL}_{n,\varepsilon,\zeta}\varphi(x_i) \leqslant -cK|x_i - y|^{-(\alpha+2-\beta)} \tag{74}$$

for all x_i with $\delta_{\varepsilon,\zeta} \leqslant |x_i - y| \leqslant c$. By the maximum principle we have

$$\min_{\Gamma} g \leqslant u \leqslant \max_{\Gamma} g.$$

Therefore, we can choose K large enough so that $\varphi(x_i) > u(x_i)$ for $|x_i - y| \geqslant c$. We trivially have $u(x_i) < \varphi(x_i)$ for $|x_i - y| \leqslant \delta_{\varepsilon,\zeta}$. Since $\mathcal{GL}_{n,\varepsilon,\zeta}(u - \varphi) \leqslant 0$ for $\delta_{\varepsilon,\zeta} \leqslant |x_i - y| \leqslant c$, the maximum principle on a graph yields $u \leqslant \varphi$ on $\mathcal{X}_n$, which completes the proof. □

We now establish a local regularity result that allows us to fill in the gap within the ball $B(x, C\varepsilon)$. The local result depends only on the variational structure of the problem, and does not use a barrier argument.

Proposition 4.8 (Local Hölder estimate) *Let $u \in L^2(\mathcal{X}_n)$. For each $z \in \Gamma$ the event that*

$$|u(x) - u(y)|^2 \leqslant \frac{Cn\varepsilon^2}{\min\{\gamma_\zeta(x), \gamma_\zeta(y)\}} \mathcal{GE}_{n,\varepsilon,\zeta}(u), \tag{75}$$

holds for all $x, y \in \mathcal{X}_n \cap B(z, r)$ with $|x - y| \leqslant \varepsilon$ occurs with probability at least $1 - C\exp\left(-cn\varepsilon^d + \log(n)\right)$.

Proof Let $z \in \Gamma$, and fix $r > 0$. Partition the cube $K := \prod_{i=1}^d [z_i - r, z_i + r]$ into hypercubes $K_1, \ldots, K_m$ of side length $h > 0$, where $m = (2r/h)^d$. Let Z_i denote the number of random variables falling in cube K_i. By Lemma 4.4 we have

$$\mathbb{P}(Z_i \leqslant \mathbb{E}[Z_i] - Cnh^d\lambda) \leqslant \exp\left(-cnh^d\lambda^2\right) \tag{76}$$

for any $0 < \lambda \leqslant 1$. Since $\mathbb{E}[Z_i] = nh^d$ we have

$$\mathbb{P}\left(\min_{1\leqslant i\leqslant m} Z_i \leqslant \tfrac{1}{2}nh^d\right) \leqslant m\exp\left(-cnh^d\right). \tag{77}$$

Let $x, y \in \mathcal{X}_n \cap B(z, r)$ such that $|x - y| \leqslant \varepsilon$, and let $\overline{x} = (x + y)/2 \in B(z, r)$. Let K_i denote the cube to which $\overline{x}$ belongs. Then for all $w \in K_i$ we have $|\overline{x} - w| \leqslant \sqrt{d}h$. Therefore, if $\sqrt{d}h \leqslant \varepsilon/2$ then

$$|x - w| \leqslant |x - \overline{x}| + |\overline{x} - w| \leqslant \frac{\varepsilon}{2} + \sqrt{d}h \leqslant \varepsilon,$$

and $|y - w| \leqslant \varepsilon$ for all $w \in K_i$. It follows that

$$K_i \subset B(x, \varepsilon) \cap B(y, \varepsilon).$$

For the remainder of the proof, we set $h = \varepsilon/(2\sqrt{d})$ and restrict ourselves to the event that

$$\min_{1\leqslant i\leqslant m} Z_i \geqslant cn\varepsilon^d \tag{78}$$

Let

$$K = \mathcal{X}_n \cap B(x, \varepsilon) \cap B(y, \varepsilon).$$

Note that for any $z \in K$ we have

$$\min\{|u(x) - u(z)|, |u(y) - u(z)|\} \geqslant \frac{1}{2}|u(x) - u(y)|.$$

Now we have

$$\begin{aligned}
&\mathcal{GE}_{n,\varepsilon,\zeta}(u) \\
&\quad = \frac{1}{2n^2\varepsilon^2} \sum_{x,y\in\mathcal{X}_n} (\gamma_\zeta(x)+\gamma_\zeta(y))\eta_\varepsilon(x-y)|u(x)-u(y)|^2 \\
&\quad \geqslant \frac{1}{2n^2\varepsilon^2} \left[\sum_{z\in\mathcal{X}_n} \gamma_\zeta(x)\eta_\varepsilon(x-z)|u(x)-u(z)|^2 + \gamma_\zeta(y)\eta_\varepsilon(y-z)|u(y)-u(z)|^2 \right] \\
&\quad \geqslant \frac{c\min\{\gamma_\zeta(x),\gamma_\zeta(y)\}|K|}{n^2\varepsilon^{d+2}}|u(x)-u(y)|^2.
\end{aligned}$$

Since $K_i \cap \mathcal{X}_n \subset K$, we have $|K| \geqslant cn\varepsilon^d$, and hence

$$|u(x)-u(y)|^2 \leqslant \frac{Cn\varepsilon^2}{\min\{\gamma_\zeta(x),\gamma_\zeta(y)\}}\mathcal{GE}_{n,\varepsilon,\zeta}(u), \tag{79}$$

which completes the proof. □

We are now equipped to give the proof of Theorem 4.1.

Proof of Theorem 4.1 The proof combines the macroscopic Hölder estimate (Proposition 4.7), and the local Hölder estimate (Proposition 4.8), and is split into two steps.

1. We note that

$$r_\zeta = \frac{r_0}{(\zeta-1)^{1/\alpha}} \leqslant C\varepsilon, \tag{80}$$

as $\zeta \geqslant 1+\varepsilon^{-\alpha}$. By (80) and Theorem 4.7, the event that

$$|u(x)-u(z)| \leqslant C|x-z|^\beta + \sup_{x\in\mathcal{X}_n\cap B(z,C\varepsilon)} |u(x)-u(y)|, \tag{81}$$

holds for all $x \in \mathcal{X}_n$ occurs with probability at least $1-C\exp(-cn\varepsilon^{d+4}+\log(n))$.

2. We note that with probability at least $1-C\exp\left(-cn\varepsilon^d+\log(n)\right)$ we have $\mathcal{GE}_{n\varepsilon,\zeta}(u) \leqslant C$ for a constant C. Therefore, by Proposition 4.8 we have that

$$|u(x)-u(y)|^2 \leqslant \frac{Cn\varepsilon^2}{\min_{B(z,C\varepsilon)}\gamma_\zeta} \tag{82}$$

holds for all $x, y \in \mathcal{X}_n \cap B(z, C\varepsilon)$ with $|x-y| \leqslant \varepsilon$ with probability at least $1-C\exp\left(-cn\varepsilon^d+\log(n)\right)$. As in the proof of Proposition 4.8, we partition the cube $K := \prod_{i=1}^d [z_i - C\varepsilon, z_i + C\varepsilon]$ into hypercubes of side length $h < \varepsilon$, and find that all cubes have at least one point from $\mathcal{X}_n$ with probability at least $1-C\exp(-cn\varepsilon^d+\log(n))$. Thus, by traversing neighboring cubes, we can construct a path from $z \in \Gamma$ to any $x \in B(z, C\varepsilon)$ consisting of at most a constant

number of points from $\mathcal{X}_n \cap B(z, C\varepsilon)$, with each step in the path smaller than ε. Applying (82) along the path yields

$$|u(x) - u(z)|^2 \leqslant \frac{Cn\varepsilon^2}{\min\{(C\varepsilon)^{-\alpha}, \zeta\}}$$

for all $x \in \mathcal{X}_n \cap B(z, C\varepsilon)$ with probability at least $1 - C\exp\left(-cn\varepsilon^d + \log(n)\right)$. Since $\zeta \geqslant \varepsilon^{-\alpha}$ we deduce

$$\sup_{x \in \mathcal{X}_n \cap B(z, C\varepsilon)} |u(x) - u(z)|^2 \leqslant Cn\varepsilon^{2+\alpha},$$

which completes the proof.

□

5 Numerical Experiments

We present the results of several numerical experiments illustrating the properly-weighted Laplacian and comparing it with the nonlocal [38] and standard graph Laplacian on real and synthetic data. All experiments were performed in Matlab and use Matlab backslash to solve the graph Laplacian system. We mention there are indirect solvers that may be faster in certain applications, such as preconditioned conjugate gradient [25], algebraic multigrid [8,25,36], or more recent fast Laplacian solvers [29,42]. Thus, the CPU times reported below have the potential to be improved substantially.

5.1 Comparison of the Profiles Obtained

First, we perform an experiment with two labels on the box $[0, 1]^d$ to illustrate our method and the differences with the nonlocal graph Laplacian [38]. The graph is a sequence of n *i.i.d.* random variables uniformly distributed on the unit box $[0, 1]^d$ in $\mathbb{R}^d$, and two labels are given $g(0, 0.5, \ldots, 0.5) = 0$ and $g(1, 0.5, \ldots, 0.5) = 1$. We set $\varepsilon = 2/n^{1/d}$ and $r_0 = 1$. The weights follow a Gaussian distribution with $\sigma = \varepsilon/2$. In Fig. 2 we show plots of the triangulated surface representing the learned function u on the graph for various values of α and ζ. Here, $n = 10^5$ and $d = 2$, and each simulation took approximately 1.5 s of CPU time. We notice that as α is increased, the learned functions are smoother in a vicinity of each label. The case of $\alpha = 0$ corresponds to the standard graph Laplacian, and returns an approximately constant label $u = 0.5$, which illustrates the degeneracy of the standard Laplacian with few labels. We note that as α is increased, we must increase ζ as well (recall (12)), otherwise the ball $B(x, r_\zeta)$ on which γ is truncated to value ζ becomes very large, and the method reduces to the standard graph Laplacian. This simply illustrates that the implicit rate in the condition $\zeta \gg n\varepsilon^2$ in Theorem 3.1 depends on α.

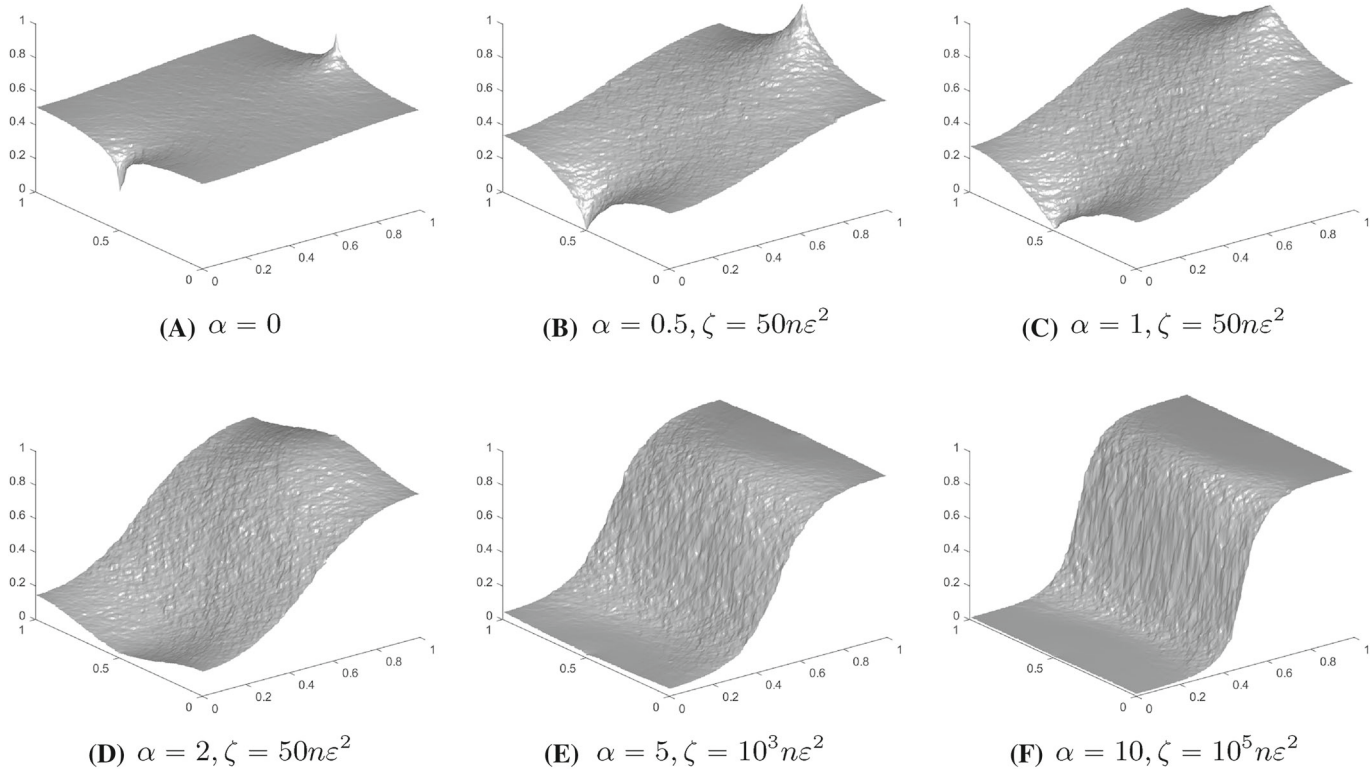
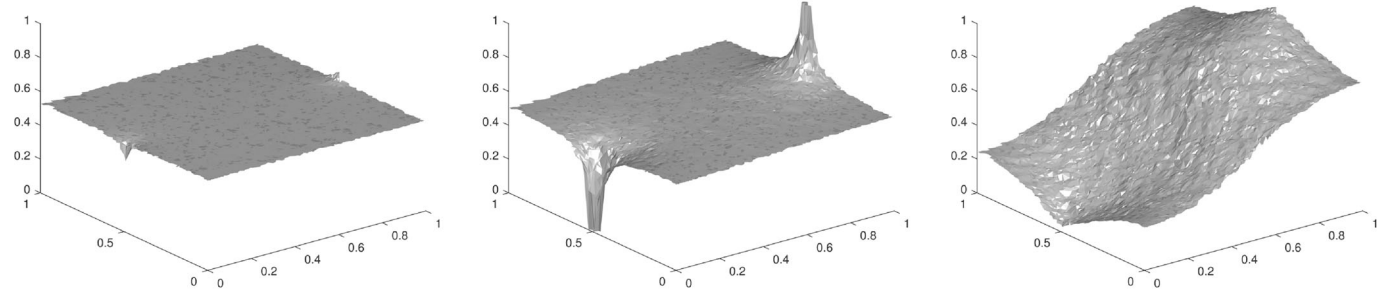


Fig. 2 Comparison of **a** the standard graph Laplacian, and **b–f** our weighted graph Laplacian with various values of α and ζ. The graph is a random geometric graph (described in the text) on $[0, 1]^2$, and the labels are $g(0, 0.5) = 0$ and $g(1, 0.5) = 1$



(A) Graph Laplacian **(B)** Weighted Lap. [38] **(C)** PW Laplacian ($\alpha = 2$)

Fig. 3 Comparison of **a** the standard graph Laplacian, **b** the nonlocal graph Laplacian [38], and **c** our properly-weighted Laplacian. The graph is a random geometric graph (described in the text) on $[0, 1]^3$, and the surfaces plotted are a slice of the learned function u near $x_3 = 0.5$. The labels are $g(0, 0.5, 0.5) = 0$ and $g(1, 0.5, 0.5) = 1$

In Fig. 3, we use the same model, but with $n = 2 \times 10^5$ points in dimension $d = 3$, and set $\zeta = 50n\varepsilon^2$. For visualization, we show the learned function restricted to a neighborhood of the slice $x_3 = 0.5$. Figure 3b illustrates the degeneracy of the nonlocal graph Laplacian [38], which returns a nearly constant label function. In contrast, our method, show in Fig. 3c, smoothly interpolates between the two labels. Each simulation in Fig. 3 took approximately 25 s to compute.

5.2 Comparison of Decision Boundaries in 2D

We now give a synthetic classification example. The graph consists of $n = 10^5$ *i.i.d* uniform random variables on $[0, 1]^2$, and the weights are chosen to be Gaussian with $\sigma = \varepsilon/2$. We chose $\alpha = 5$, $\zeta = 10^6 n\varepsilon^2$, $r_0 = 1$, and $\varepsilon = 3/\sqrt{n}$. Two labels, $g(0, 0) = 0$ and $g(1, 1) = 1$ are provided. Figure 4 shows the decision boundaries (i.e., the level-set $\{u = 0.5\}$) over 25 trials for the standard graph Laplacian, the nonlocal Laplacian, and our method. Each trial took roughly 1.5 s to compute. We see that the nonlocal and standard Laplacian are highly sensitive to small variations in the graph, giving a wide variety of results over the 25 trials. This is a reflection of the degeneracy, or ill-posedness, in the small label regime, and suggests the methods are very sensitive to perturbations in the data. In contrast, our method very consistently divides the square along the diagonal.

5.3 Comparison of Classes Obtained in 3D

We consider samples of the measure supported on domain is $[0, 1]^3$ and with density 1 except for the strip $[0.45, 0.55] \times [0, 1] \times [0, 1]$ where the density is 0.6. We considered 20 runs with $n = 50,000$ points in the domain. The given labeled points are $g(0, 0.2, 0.2) = 0$ and $g(1, 0.2, 0.2) = 1$. Due to the symmetry, the correct decision boundary is the plane $x_1 = 0.5$. We used a connectivity distance for the graph construction of $\varepsilon = 3/n^{\frac{1}{3}}$, which yielded a typical vertex degree of about 116. We consider Gaussian weights with $\sigma = \varepsilon/2$. We chose $\alpha = 5$, $\zeta = 10^6 n\varepsilon^2$, $r_0 = 1$ for our method. A typical result for one run is illustrated in Fig. 5. The standard graph Laplacian produced very unstable results with the average of 49.8% misclassified points. The nonlocal Laplacian of [38] was also rather unstable with sometimes almost perfect decision boundary and sometimes large sections of misclassified points. On average it misclassified 11% of points. Our method was stable and in all experiments identified the correct boundary, with average classification error of 0.25%. We observed similar outcomes for a variety of sets of parameters.

5.4 Comparison on the MNIST Dataset

Our last experiment considers classification of handwritten digits from the MNIST dataset, which consists of 70,000 grayscale 28×28 pixel images of handwritten digits 0–9 [30]. Figure 6 shows examples of some of the images in the MNIST dataset. MNIST is estimated to have intrinsic dimension between $d = 12$ and $d = 14$ [14,28],

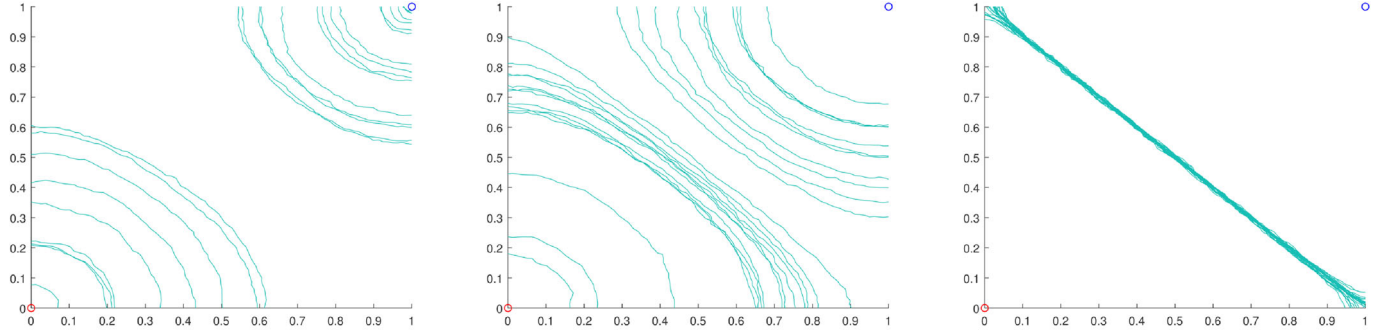
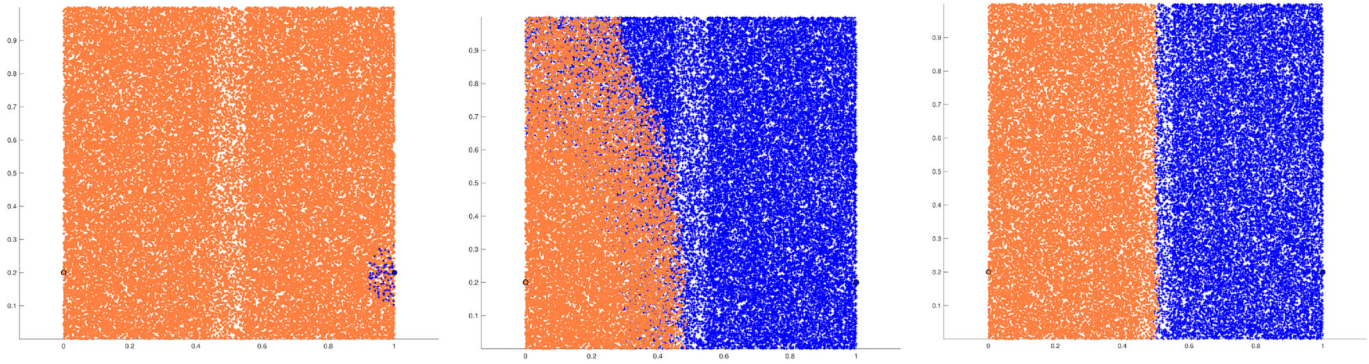


Fig. 4 Comparison of decision boundaries plotted over 25 random trials for a synthetic classification problem for **a** the standard graph Laplacian, **b** the nonlocal graph Laplacian [38], and **c** our properly-weighted graph Laplacian. Two labels are given, $g(0, 0) = 0$ (red point) and $g(1, 1) = 1$ (blue point), and the graph is a uniform random geometric graph (described in the text) (Color figure online)



(A) Graph Laplacian **(B)** Weighted Lap. [38] **(C)** Our method ($\alpha = 5$)

Fig. 5 Comparison of results for a synthetic classification problem for **a** the standard graph Laplacian, **b** the nonlocal graph Laplacian [38], and **c** our weighted graph Laplacian. The domain is $[0, 1]^3$ and the density is 1 except for the strip $[0.45, 0.55] \times [0, 1] \times [0, 1]$ where it is 0.6. The given labeled points are $g(0, 0.2, 0.2) = 0$ and $g(1, 0.2, 0.2) = 1$. There are $n = 50,000$ points in the domain. Connectivity distance for the graph construction is $3/n^{\frac{1}{3}}$ and for our method $\alpha = 5$

Fig. 6 Example of some of the handwritten digits from the MNIST dataset [30]

Table 1 Accuracy for classification of MNIST handwritten digits with 10, 30, 50, 70 and 100 labels via the standard graph Laplacian, the nonlocal weighted Laplacian [38], and our properly-weighted Laplacian

# Labels	10		30		50		70		100	
Method	Mean (%)	Std	Mean (%)	Std	Mean (%)	Std	Mean (%)	Std	Mean (%)	Std
Graph laplacian	14.2	6.3	24.3	11.9	53	10.9	68	6.4	76.1	7.6
Weighted lap. [38]	67.9	8.7	84.8	2.7	88.8	1.1	89.6	1.3	90.9	1.1
PW-laplacian	68	8.6	84.9	2.7	88.8	1.1	89.6	1.3	90.9	1.1

The results are averaged over 10 trials, and the mean and standard deviation of accuracy are reported

which suggests a larger value for α is appropriate. We used all 70, 000 MNIST images to construct the graph. Our construction is the same as in [38]; we connect each data point to its nearest 50 neighbors (in Euclidean distance), and assign Gaussian weights taking σ to be the distance to the 20th nearest neighbor. We symmetrize the graph by replacing the weight matrix W with $\frac{1}{2}(W^T + W)$, which is done automatically by the variational formulation (recall (19)). We then take L randomly chosen images from each class (digit) as labels, where $L = 1, 3, 5, 7, 10$, and provide the true labels for these digits. The semi-supervised learning algorithm performs 10 binary classifications, for each digit versus the rest, which generates functions $u^0, u^1, u^2, \ldots, u^9$ on the graph. The label for each image x in the dataset is chosen as the index i for which $u^i(x)$ is maximal. The algorithm is standard in semi-supervised learning, and identical to the one used in [38].

For each value of $L \in \{1, 3, 5, 7, 10\}$, we ran the experiment described above 10 times, choosing the labels randomly (in the same way for each algorithm) every time. Each of the 500 trials took approximately 15 min to compute in Matlab. The mean and standard deviation of accuracy are shown in Table 1. Our method performs very similarly to the nonlocal Laplacian [38], and both significantly outperform the standard graph Laplacian. We ran our method for $\alpha = 2, 5, 10$, producing nearly identical results in all cases. We used $\zeta = 10^7$ and $r_0 = 0.1$. We found the results for our method were largely insensitive to many of the parameters in our algorithm; the accuracy begins to decrease when $\alpha < 1$ and when $r_0 > 1$. We note that the accuracy scores reported in Table 1 are much higher than those reported in [38]; we believe this is because the authors in [38] subsampled MNIST to 16,000 images. This observation speaks favorably to the semi-supervised paradigm that learning can be improved by access to additional unlabeled data. We note that the accuracy scores

for our method and the nonlocal weighted Laplacian [38] are identical (to one significant digit) for 30, 50, 70, and 100 labels. Most data points in MNIST are relatively far from their nearest neighbors, and so our nonlocal weights have less effect, compared to the low dimensional examples presented above. For this reason, the weight matrix for our method is very similar to the nonlocal Laplacian [38]. We expect to see more of a difference in applications to larger datasets. For example, it would be interesting (and challenging) to apply these techniques to a dataset like ImageNet [16], which consists of over 14 million natural images belonging to over 20,000 categories.

Acknowledgements Calder was supported by NSF Grant DMS:1713691. Slepčev acknowledges the NSF support (Grants DMS-1516677 and DMS-1814991). He is grateful to University of Minnesota, where this project started, for hospitality. He is also grateful to the Center for Nonlinear Analysis of CMU for its support. Both authors thank the referees for valuable suggestions.

Appendix A: Background Material

Here we recall some of the notions our work depends on and establish an auxiliary technical result.

A.1 Γ–Convergence

Γ-convergence was introduced by De Giorgi in 1970s to study limits of variational problems. We refer to [7,15] for comprehensive introduction to Γ-convergence. We now recall the notion of Γ-convergence is in a random setting.

Definition A.1 (*Γ-convergence*) Let (Z, d) be a metric space, $L^0(Z; \mathbb{R} \cup \{\pm\infty\})$ be the set of measurable functions from Z to $\mathbb{R} \cup \{\pm\infty\}$, and $(\mathcal{X}, \mathbb{P})$ be a probability space. The function $\mathcal{X} \ni \omega \mapsto \mathcal{E}_n^{(\omega)} \in L^0(Z; \mathbb{R} \cup \{\pm\infty\})$ is a random variable. We say $\mathcal{E}_n^{(\omega)}$ Γ-*converge almost surely* on the domain Z to $\mathcal{E}_\infty : Z \to \mathbb{R} \cup \{\pm\infty\}$ with respect to d, and write $\mathcal{E}_\infty = 0\text{-}\lim_{n\to\infty} \mathcal{E}_n^{(\omega)}$, if there exists a set $\mathcal{X}' \subset \mathcal{X}$ with $\mathbb{P}(\mathcal{X}') = 1$, such that for all $\omega \in \mathcal{X}'$ and all $f \in Z$:

(i) (liminf inequality) for every sequence $\{u_n\}_{n=1,\dots}$ in Z converging to f

$$\mathcal{E}_\infty(f) \leqslant \liminf_{n\to\infty} \mathcal{E}_n^{(\omega)}(u_n), \text{ and}$$

(ii) (recovery sequence) there exists a sequence $\{u_n\}_{n=1,2,\dots}$ in Z converging to f such that

$$\mathcal{E}_\infty(f) \geqslant \limsup_{n\to\infty} \mathcal{E}_n^{(\omega)}(u_n).$$

For simplicity we suppress the dependence of ω in writing our functionals. The almost sure nature of the convergence in our claims in ensured by considering the set of realizations of $\{x_i\}_{i=1,\dots}$ such that the conclusions of Theorem A.3 hold (which they do almost surely).

An important result concerning Γ-convergence is that any subsequential limit of the sequence of minimizers of $\mathcal{E}_n$ is a minimizer of the limiting functional $\mathcal{E}_\infty$. So to show that the minimizers of $\mathcal{E}_n$ converge at least along a subsequence to a minimizer of $\mathcal{E}_\infty$it suffices to establish the precompactness of the set of minimizers. We make this precise in the theorem below. Its proof can be found in [7, Theorem 1.21] or [15, Theorem 7.23].

Theorem A.2 (Convergence of Minimizers) *Let (Z, d) be a metric space and $\mathcal{E}_n : Z \to [0, \infty]$ be a sequence of functionals. Let u_n be a minimizing sequence for $\mathcal{E}_n$. If the set $\{u_n\}_{n=1,2,\dots}$ is precompact and $\mathcal{E}_\infty = 0\text{-}\lim_n \mathcal{E}_n$ where $\mathcal{E}_\infty : Z \to [0, \infty]$ is not identically ∞ then*

$$\min_Z \mathcal{E}_\infty = \lim_{n\to\infty} \inf_Z \mathcal{E}_n.$$

Furthermore any cluster point of $\{u_n\}_{n=1,2,\dots}$ is a minimizer of E_∞.

The theorem is also true if we replace minimizers with approximate minimizers.

We note that Γ-convergence is defined for functionals on a common metric space. Section A.3 overviews the metric space we use to analyze the asymptotics of our semi-supervised learning models, in particular it allows us to go from discrete to continuum.

A.2 Optimal Transportation and Approximation of Measures

Here we recall the notion of optimal transportation between measures and the metric it introduces. Comprehensive treatment of the topic can be found in books of Villani [45] and Santambrogio [37].

Given a bounded, open set $\Omega \subset \mathbb{R}^d$, and probability measures μ and ν in $\mathcal{P}(\overline{\Omega})$ we define the set $\Pi(\mu, \nu)$ of transportation plans, or couplings, between μ and ν to be the set of probability measures on the product space $\pi \in \mathcal{P}(\overline{\Omega} \times \overline{\Omega})$ whose first marginal is μ and second marginal is ν. We then define the p-optimal transportation distance (a.k.a. p-Wasserstein distance) by

$$d_p(\mu, \nu) = \begin{cases} \inf\limits_{\pi\in\Pi(\mu,\nu)} \left(\int_{\Omega\times\Omega} |x - y|^p \, d\pi(x, y) \right)^{\frac{1}{p}} & \text{if } 1 \leqslant p < \infty \\ \inf\limits_{\pi\in\Pi(\mu,\nu)} \pi\text{-esssup}_{(x,y)} |x - y| & \text{if } p = \infty. \end{cases}$$

If μ is absolutely continuous with respect to the Lebesgue measure on Ω, then the distance can be rewritten using transportation maps, $T : \Omega \to \Omega$, instead of transportation plans,

$$d_p(\mu, \nu) = \begin{cases} \inf\limits_{T_\#\mu=\nu} \left(\int_{\Omega} |x - T(x)|^p \, d\mu(x) \right)^{\frac{1}{p}} & \text{if } 1 \leqslant p < \infty \\ \inf\limits_{T_\#\mu=\nu} \mu\text{-esssup}_x |x - T(x)| & \text{if } p = \infty. \end{cases}$$

where $T_{\#}\mu = \nu$ means that the push forward of the measure μ by T is the measure ν, namely that T is Borel measurable and such that for all $U \subset \overline{\Omega}$, open, $\mu(T^{-1}(U)) = \nu(U)$.

When $p < \infty$ the metric d_p metrizes the weak* convergence of measures.

Optimal transportation plays an important role in comparing the discrete and continuum objects we study. In particular, we use sharp estimates on the ∞-optimal transportation distance between a measure and the empirical measure of its sample. In the form below, for $d \geqslant 2$, they were established in [19], which extended the related results in [1,31,39,43].

Theorem A.3 *For $d \geqslant 2$, let $\Omega \subset \mathbb{R}^d$ be open, connected and bounded with Lipschitz boundary. Let μ be a probability measure on Ω with density (with respect to Lebesgue) ρ which is bounded above and below by positive constants. Let $x_1, x_2, \ldots$ be a sequence of independent random variables with distribution μ and let μ_n be the empirical measure. Then, there exists constants $C \geqslant c > 0$ such that almost surely there exists a sequence of transportation maps $\{T_n\}_{n=1}^{\infty}$ from μ to μ_n with the property*

$$c \leqslant \liminf_{n\to\infty} \frac{\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}}{\ell_n} \leqslant \limsup_{n\to\infty} \frac{\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}}{\ell_n} \leqslant C$$

where

$$\ell_n = \begin{cases} \frac{(\ln n)^{\frac{3}{4}}}{\sqrt{n}} & \text{if } d = 2 \\ \frac{(\ln n)^{\frac{1}{d}}}{n^{\frac{1}{d}}} & \text{if } d \geqslant 3. \end{cases} \tag{83}$$

A.3 The TL^p Space

The discrete functionals we consider (e.g. $\mathcal{E}_{n,\varepsilon_n,\zeta_n}$) are defined for functions $u_n : \mathcal{X}_n \to \mathbb{R}$, while the limit functional $\mathcal{E}$ acts on functions $f : \Omega \to \mathbb{R}$, where Ω is an open set. We can view u_n as elements of $L^p(\mu_n)$ where μ_n is the empirical measure of the sample $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. Likewise $f \in L^p(\mu)$ where μ is the measure with density ρ from which the data points are sampled. In order to compare f and u_n in a way that is consistent with the L^p topology we use the TL^p space that was introduced in [20], where it was used to study the continuum limit of the graph total variation. Subsequent development of the TL^p space has been carried out in [21,44].

To compare the functions u_n and f above we need to take into account their domains, or more precisely to account for μ and μ_n. For that purpose the space of configurations is defined to be

$$TL^p(\Omega) = \left\{(\mu, f) \,:\, \mu \in \mathcal{P}(\overline{\Omega}),\, f \in L^p(\mu)\right\}.$$

The metric on the space is

$$d^p_{TL^p}((\mu, f), (\nu, g))$$

$$= \inf \left\{ \int_{\Omega\times\Omega} |x-y|^p + |f(x)-g(y)|^p \, d\pi(x,y) \, : \, \pi \in \Pi(\mu,\nu) \right\}$$

where $\Pi(\mu,\nu)$ the set of transportation plans defined in Sect. A.2. We note that the minimizing π exists and that TL^p space is a metric space, [20].

As shown in [20], when μ is absolutely continuous with respect to the Lebesgue measure on Ω, then the distance can be rewritten using transportation maps T, instead of transportation plans,

$$\begin{aligned} &d^p_{TL^p}((\mu, f), (\nu, g)) \\ &\quad = \inf \left\{ \int_{\Omega} |x - T(x)|^p + |f(x) - g(T(x))|^p \, d\mu(x) \, : \, T_{\#}\mu = \nu \right\} \end{aligned}$$

where the push forward of the measure $T_{\#}\mu$ is defined in Sect. A.2. This formula provides an interpretation of the distance in our setting. Namely, to compare functions $u_n : \mathcal{X}_n \to \mathbb{R}$ we define a mapping $T_n : \Omega \to \mathcal{X}_n$ and compare the functions $\widetilde{f_n} = u_n \circ T_n$ and f in $L^p(\mu)$, while also accounting for the transport, namely the $|x - T_n(x)|^p$ term.

We remark that the $TL^p(\overline{\Omega})$ space is not complete, and that its completion was discussed in [20]. In the setting of this paper, since the corresponding measure is clear from context, we often say that u_n converges in TL^p to f as a short way to say that (μ_n, u_n) converges in TL^p to (μ, f).

A.4 Local Estimates for Weighted Laplacian

Lemma A.4 *There exists $C > 0$ such that for each $u \in H^1(B(0,1))$ there exists $v \in H^1(B(0,1))$ such that*

$$\begin{aligned} v|_{B(0,\frac{1}{2})} &\equiv \frac{1}{|B(0,\frac{1}{2})|} \int_{B(0,\frac{1}{2})} u(x)dx \\ v|_{\partial B(0,1)} &= u|_{\partial B(0,1)} \\ \|\nabla v\|_{L^2(B(0,1))} &\leqslant C\|\nabla u\|_{L^2(B(0,1))} \end{aligned}$$

where the value on the boundary is considered in sense of the $L^2(\partial B(0,1))$ trace.

Proof Let

$$\overline{u} = \frac{1}{|B(0,\frac{1}{2})|} \int_{B(0,\frac{1}{2})} u(x)dx.$$

Let $\phi \in C^\infty([0,1],[0,1])$ be such that $\phi(r) = 1$ for all $r \in [0, \frac{1}{2}]$ and $\phi(1) = 0$. Let $M = \max_{r\in[0,1]} |\phi'(r)|$. Let

$$v(x) = \phi(|x|)\overline{u} + (1 - \phi(|x|))u(x).$$

By Poincaré inequality stated in Theorem 13.27 of [32] there exists $C_1 > 0$, independent of u,

$$\int_{B(0,1)} |u(x) - \overline{u}|^2 dx \leqslant C_1 \int_{B(0,1)} |\nabla u(x)|^2 dx.$$

Using the Poincaré inequality we obtain, for $C = 2 + 2MC_1$,

$$\int_{B(0,1)} |\nabla v|^2 dx \leqslant 2 \int_{B(0,1)} |\nabla u|^2 + M|u - \overline{u}|^2 dx \leqslant C \int_{B(0,1)} |\nabla u|^2 dx.$$

□

References

1. Ajtai, M., Komlós, J., Tusnády, G.: On optimal matchings. Combinatorica **4**(4), 259–264 (1984)
2. Alamgir, M., Luxburg, U.V.: Phase transition in the family of p-resistances. In: Advances in Neural Information Processing Systems, pp. 379–387 (2011)
3. Ando, R.K., Zhang, T.: Learning on graph with laplacian regularization. Adv. Neural Inf. Process. Syst. **19**, 25 (2007)
4. Bardi, M., Capuzzo-Dolcetta, I.: Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations. Springer, New York (2008)
5. Belkin, M., Niyogi, P.: Using manifold structure for partially labeled classification. In: Advances in Neural Information Processing Systems (NIPS), pp. 953–960 (2003)
6. Bertozzi, A., Luo, X., Stuart, A., Zygalakis, K.: Uncertainty quantification in graph-based classification of high dimensional data. SIAM/ASA J. Uncertain. Quantif. **6**(2), 568–595 (2018)
7. Braides, A.: Γ-Convergence for Beginners. Oxford Lecture Series in Mathematics and Its Applications, vol. 22. Oxford University Press, Oxford (2002)
8. Brandt, A., Livne, O.E.: Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamics, vol. 67. SIAM, Philadelphia (2011)
9. Bridle, N., Zhu, X.: p-voltages: laplacian regularization for semi-supervised learning on high-dimensional data. In: Eleventh Workshop on Mining and Learning with Graphs (MLG2013) (2013)
10. Calder, J.: The game theoretic p-Laplacian and semi-supervised learning with few labels. Nonlinearity **32**(1), 301–330 (2018)
11. Calder, J.: Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data. To appear in SIAM Journal on Mathematics of Data Science (2019)
12. Calder, J., Smart, C.K.: The limit shape of convex hull peeling. arXiv preprint arXiv:1805.08278 (2018)
13. Chapelle, O., Scholkopf, B., Zien, A.: Semi-supervised Learning. MIT, London (2006)
14. Costa, J.A., Hero, A.O.: Determining intrinsic dimension and entropy of high-dimensional shape spaces. In: Statistics and Analysis of Shapes, pp. 231–252. Springer, New York (2006)
15. Dal Maso, G.: An introduction to Γ-Convergence. Progress in Nonlinear Differential Equations and Their Applications, vol. 8. Birkhäuser Boston Inc, Boston (1993)
16. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009., pp. 248–255. IEEE (2009)
17. Dunlop, M.M., Slepčev, D., Stuart, A.M., Thorpe, M.: Large data and zero noise limits of graph-based semi-supervised learning algorithms. Appl. Comput. Harmon. Anal. (2019). https://doi.org/10.1016/j.acha.2019.03.005
18. El Alaoui, A., Cheng, X., Ramdas, A., Wainwright, M.J., Jordan, M.I.: Asymptotic behavior of lp-based Laplacian regularization in semi-supervised learning. In: 29th Annual Conference on Learning Theory, pp. 879–906 (2016)
19. García Trillos, N., Slepčev, D.: On the rate of convergence of empirical measures in ∞-transportation distance. Can. J. Math. **67**(6), 1358–1383 (2015)

20. García Trillos, N., Slepčev, D.: Continuum limit of total variation on point clouds. Arch. Ration. Mech. Anal. **220**(1), 193–241 (2016)
21. García Trillos, N., Slepčev, D.: A variational approach to the consistency of spectral clustering. Appl. Comput. Harmon. Anal. **45**(2), 239–281 (2018)
22. García Trillos, N., Slepčev, D., von Brecht, J., Laurent, T., Bresson, X.: Consistency of Cheeger and ratio graph cuts. J. Mach. Learn. Res. **17**(1), 6268–6313 (2016)
23. García Trillos, N., Gerlach, M., Hein, M., Slepčev, D.: Error estimates for spectral convergence of the graph Laplacian on random geometric graphs towards the Laplace-Beltrami operator. Found. Comput. Math. (2019). https://doi.org/10.1007/s10208-019-09436-w
24. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order. Springer, New York (2015)
25. Greenbaum, A.: Iterative Methods for Solving Linear Systems, vol. 17. SIAM, Philadelphia (1997)
26. He, J., Li, M., Zhang, H.-J., Tong, H., Zhang, C.: Manifold-ranking based image retrieval. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 9–16. ACM (2004)
27. He, J., Li, M., Zhang, H.-J., Tong, H., Zhang, C.: Generalized manifold-ranking-based image retrieval. IEEE Trans. Image Process. **15**(10), 3170–3177 (2006)
28. Hein, M., Audibert, J.-Y.: Intrinsic dimensionality estimation of submanifolds in Rd. In: Proceedings of the 22nd International Conference on Machine learning, pp. 289–296. ACM (2005)
29. Kyng, R., Rao, A., Sachdeva, S., Spielman, D.A.: Algorithms for lipschitz learning on graphs. In: Proceedings of The 28th Conference on Learning Theory, pp. 1190–1223 (2015)
30. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
31. Leighton, T., Shor, P.: Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. Combinatorica **9**(2), 161–187 (1989)
32. Leoni, G.: A First Course in Sobolev Spaces, vol. 181. American Mathematical Society, Providence (2017)
33. Luxburg, U.V., Bousquet, O.: Distance-based classification with lipschitz functions. J. Mach. Learn. Res. **5**(Jun), 669–695 (2004)
34. Nadler, B., Srebro, N., Zhou, X.: Semi-supervised learning with the graph Laplacian: the limit of infinite unlabelled data. In: Neural Information Processing Systems (NIPS) (2009)
35. Rudin, W.: Real and Complex Analysis. Tata McGraw-Hill Education, New York (2006)
36. Ruge, J.W., Stüben, K.: Algebraic multigrid. In: Multigrid Methods, pp. 73–130. SIAM (1987)
37. Santambrogio, F.: Optimal Transport for Applied Mathematicians, Progress in Nonlinear Differential Equations and their Applications Calculus of variations, PDEs, and modeling, vol. 87. Birkhäuser, Cham (2015)
38. Shi, Z., Osher, S., Zhu, W.: Weighted nonlocal laplacian on interpolation from sparse data. J. Sci. Comput. **73**(2–3), 1164–1177 (2017)
39. Shor, P.W., Yukich, J.E.: Minimax grid matching and empirical measures. Ann. Probab. **19**(3), 1338–1348 (1991)
40. Singer, A.: From graph to manifold Laplacian: the convergence rate. Appl. Comput. Harmon. Anal. **21**(1), 128–134 (2006)
41. Slepčev, D., Thorpe, M.: Analysis of p-Laplacian regularization in semi-supervised learning. SIAM J. Math. Anal. **51**(3), 2085–2120 (2019)
42. Spielman, D.A., Teng, S.-H.: Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In: Proceedings of the Thirty-Sixth Annual ACM Symposium on Theory of Computing, pp. 81–90. ACM (2004)
43. Talagrand, M.: Upper and Lower Bounds of Stochastic Processes. Modern Surveys in Mathematics, vol. 60. Springer, Berlin (2014)
44. Thorpe, M., Park, S., Kolouri, S., Rohde, G.K., Slepčev, D.: A transportation L^p distance for signal analysis. J. Math. Imaging Vis. **59**(2), 187–210 (2017)
45. Villani, C.: Optimal transport, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338. Springer, Berlin (2009)
46. Wang, Y., Cheema, M.A., Lin, X., Zhang, Q.: Multi-manifold ranking: Using multiple features for better image retrieval. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 449–460. Springer, New York (2013)

47. Xu, B., Bu, J., Chen, C., Cai, D., He, X., Liu, W., Luo, J.: Efficient manifold ranking for image retrieval. In: Proceedings of the 34th international ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 525–534. ACM (2011)
48. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.-H.: Saliency detection via graph-based manifold ranking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3166–3173 (2013)
49. Zhou, X., Belkin, M.: Semi-supervised learning by higher order regularization. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, volume 15 of Proceedings of Machine Learning Research, pp. 892–900 (2011)
50. Zhou, D., Schölkopf, B.: Regularization on discrete spaces. In: Proceedings of the 27th DAGM Conference on Pattern Recognition, PR'05, pp. 361–368. Springer, Berlin (2005)
51. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schölkopf, B.: Learning with local and global consistency. Adv. Neural Inf. Process. Syst. **16**(16), 321–328 (2004)
52. Zhou, D., Weston, J., Gretton, A., Bousquet, O., Schölkopf, B.: Ranking on data manifolds. Adv. Neural Inf. Process. Syst. **16**, 169–176 (2004)
53. Zhou, D., Huang, J., Schölkopf, B.: Learning from labeled and unlabeled data on a directed graph. In: Proceedings of the 22nd International Conference on Machine Learning, pp. 1036–1043. ACM (2005)
54. Zhou, X., Belkin, M., Srebro, N.: An iterated graph laplacian approach for ranking on manifolds. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 877–885. ACM (2011)
55. Zhu, X., Ghahramani, Z., Lafferty, J., et al.: Semi-supervised learning using Gaussian fields and harmonic functions. Int. Conf. Mach. Learn. **3**, 912–919 (2003)