Analysis of p-Laplacian Regularization in Semi-Supervised Learning

Dejan Slepčev[1] and Matthew Thorpe[1]

[1]Carnegie Mellon University, Pittsburgh, PA 15213, USA

19th July 2017

Abstract

We investigate a family of regression problems in a semi-supervised setting. The task is to assign real-valued labels to a set of n sample points, provided a small training subset of N labeled points. A goal of semi-supervised learning is to take advantage of the (geometric) structure provided by the large number of unlabeled data when assigning labels. We consider a random geometric graph, with connection radius $\varepsilon(n)$, to represent the geometry of the data set. We study objective functions which reward the regularity of the estimator function and impose or reward the agreement with the training data. In particular we consider discrete p-Laplacian regularization.

We investigate asymptotic behavior in the limit where the number of unlabeled points increases while the number of training points remains fixed. We uncover a delicate interplay between the regularizing nature of the functionals considered and the nonlocality inherent to the graph constructions. We rigorously obtain almost optimal ranges on the scaling of $\varepsilon(n)$ for the asymptotic consistency to hold. We discover that for standard approaches used thus far there is a restrictive upper bound on how quickly $\varepsilon(n)$ must converge to zero as $n \to \infty$. Furthermore we introduce a new model which overcomes this restriction. It is as simple as the standard models, but converges as soon as $\varepsilon(n) \to 0$ as $n \to \infty$.

1 Introduction

Due to its applicability across a large spectrum of problems semi-supervised learning (SSL) is a important tool in data analysis. It deals with situations when one has access to relatively few labeled points but potentially a large number of unlabeled data. We assume that we are given N labeled points $\{(x_i, y_i) \,:\, i = 1, \dots, N,\ x_i \in \mathbb{R}^d,\ y_i \in \mathbb{R}\}$ and $n - N$ points $x_i,\ i = N+1, \dots, n$ drawn from a fixed, but unknown measure, μ supported in a compact subset of $\mathbb{R}^d$. The goal is to assign labels to the unlabeled points, while taking advantage of the information provided by the unlabeled points when designing the estimator. In particular the unlabeled points carry information on the structure of μ, such as the geometry of its support, which can lead to better estimators. To access the information on μ in a way that carries over to high dimensions, we build a graph whose vertices are data points and connect them if they are close enough, that is if they are within some distance $\varepsilon > 0$. More generally the edge weights are prescribed by using a decreasing function $\eta : [0, \infty) \to [0, \infty)$ with $\lim_{r \to \infty} \eta(r) = 0$. For fixed scale $\varepsilon > 0$ we set the weights to be

$$W_{ij} = \eta_\varepsilon(|x_i - x_j|)$$

where $\eta_\varepsilon = \frac{1}{\varepsilon^d}\eta(\cdot/\varepsilon)$.

The regression problem is to find an estimator $u : \Omega_n := \{x_i \,:\, i = 1, \dots, n\} \to \mathbb{R}$ which agrees with preassigned labels. To solve the regression problem one considers objective functions which penalize the lack of smoothness of u and take the structure of the graph into account. In particular here we consider the functionals which generalize the graph Laplacian, namely the graph p-Laplacian. A particular objective function we consider is

$$\mathcal{E}_n^{(p)}(f) = \frac{1}{\varepsilon_n^p n^2} \sum_{i,j=1}^{n} W_{ij} |f(x_i) - f(x_j)|^p$$

with the constraint that $f(x_i) = y_i$ for all $i = 1, \dots, n$.

We consider the asymptotic behavior in the limit when the number of unlabeled data goes to infinity, $n \to \infty$, which is the limit relevant to semi-supervised learning. As $n \to \infty$, $\varepsilon(n) \to 0$ to increase the resolution and limit the computational cost. Namely as $\varepsilon(n)$ is the length scale over which the information on μ is averaged, taking $\varepsilon(n)$ to zero insures that the finer scales of μ are resolved as more data points become available.

To describe the limiting problem we assume that μ has density ρ which is positive and bounded from below on an open set Ω and is zero otherwise. The limiting problem corresponds to minimizing

$$\mathcal{E}_\infty^{(p)}(f) = \sigma \int_\Omega |\nabla f|^p \, \rho^2(x) \, \mathrm{d}x$$

where σ is a constant that depends on η, subject to constraint that $f(x_i) = y_i$ for $i = 1, \dots, N$.

Finiteness of $\mathcal{E}_\infty^{(p)}(f)$ implies that f lies in the Sobolev space $W^{1,p}(\Omega)$. For the constraints to make sense it is needed that pointwise evaluation of functions is well defined, which is the case only if $p > d$, when the Sobolev embedding ensures that functions in $W^{1,p}$ are continuous. Indeed it was observed [39] that graph Laplacian based regularizations, which correspond to $p = 2$ develop spikes as $n \to \infty$. The explanation for the appearance of spikes based on the regularity of f which stems from boundedness of $\mathcal{E}_\infty^{(p)}(f)$ was provided by [14]. They identify $p = d$ as the transition point: they argue that for $p \leq d$ the minimizers of $\mathcal{E}_n^{(p)}(f)$ can develop spikes as $n \to \infty$, while for $p > d$ they should not develop spikes (the authors consider $p \geq d + 1$, but the same argument applies for $p > d$). The authors also argue that for data purposes taking $p > d$ and close to d is optimal since since as $p \to \infty$ the solution forgets the information provided by the unlabeled points and only depends on the labeled ones.

Our initial goal was to validate the conclusions of [14] and rigorously show that minimizers of $\mathcal{E}_n^{(p)}(f)$ constrained to agree with provided labels converge, in appropriate topology, to minimizers of $\mathcal{E}_\infty^{(p)}(f)$ which also respect the labels and $n \to \infty$ and $\varepsilon_n \to 0$ if and only if $p > d$. However we discovered an additional phenomenon, namely that the undesirable spikes in the minimizers to graph p-Laplacian occur even when $p > d$.

Namely [14] shows pointwise convergence of the form $\lim_{\varepsilon \to 0} \lim_{n \to \infty} \mathcal{E}_n^{(p)}(f) \to \mathcal{E}_\infty^{(p)}(f)$, when f is smooth enough. However this is not sufficient to conclude that the constrained minimizers of $\mathcal{E}_n^{(p)}$ converge to constrained minimizers of $\mathcal{E}_\infty^{(p)}$. We show, roughly speaking, that for $d \geq 3$ the convergence of minimizers holds if and only if

$$\left(\frac{1}{n}\right)^{\frac{1}{p}} \gg \varepsilon_n \gg \left(\frac{\log n}{n}\right)^{\frac{1}{d}} \qquad \text{as } n \to \infty.$$

The lower bound above is related to the connectivity of the graph constructed and was well understood. Our lower bounds for $d = 1, 2$ contain additional correction terms and are not believed to be optimal. Our upper bound implies that the algorithms used are in fact not consistent for a large family of scalings of ε on n that were thus far thought to ensure consistency (namely for $1 \gg \varepsilon_n \gg n^{-1/p}$). Our work indicates that careful analytical approaches are needed and are in fact capable of providing precise information on asymptotic consistency of algorithms.

When $\varepsilon_n^p n \to \infty$, under the usual connectivity requirement (which when $d \geq 3$ reads $\varepsilon_n^d \frac{n}{\log n} \to \infty$), we are still able to establish the asymptotic behavior of algorithms. Namely we show that minimizers of $\mathcal{E}_n^{(p)}(f)$ with constraints converge, along subsequences, as $n \to \infty$ and $\varepsilon_n \to 0$ to a minimizer of $\mathcal{E}_\infty^{(p)}(f)$ without constraints. That is, the labels are forgotten in the limit as $n \to \infty$. This explains why, for large n, minimizers of $\mathcal{E}_n^{(p)}$ are 'spikey'. The need to consider subsequences in the limit is due to the fact that minimizers of $\mathcal{E}_\infty^{(p)}(f)$ without constraints are nonunique.

While the degeneracy of the problem when $p \leq d$ was known, [14], we believe that degeneracy when $p > d$ and $\varepsilon_n^p n \to \infty$ is a new and at first surprising result. The heuristic explanation for the appearance of spikes is that the discrete p-Laplacian does not share the regularizing properties of the continuum p-Laplacian. Namely the discrete p-Laplacian still involves averaging over the length scale ε and thus more closely resembles an integral operator. This allows high-frequency irregularities to form, without paying a high price in the energy. In particular, if we consider one labeled point taking the value 1, say $f_n(x_1) = 1$, while $f_n(x_i) = 0$ for all $i \geq 2$ then

$$\mathcal{E}_n^{(p)}(f_n) = \frac{2}{\varepsilon_n^p n^2} \sum_{j=2}^{n} \frac{1}{\varepsilon_n^d} \eta \left(\frac{|x_1 - x_j|}{\varepsilon_n} \right) = \frac{2}{\varepsilon_n^p n} \, \eta_{\varepsilon_n} * \mu_n(x_1) \to 0$$

as $n \to \infty$, when $\varepsilon_n^p n \to \infty$. Note that f_n exhibits degeneracy while $\mathcal{E}_n^{(p)}(f_n) \to 0$.

In addition to the constrained problem above we also consider the problem where the agreement with the labels provided is imposed in a softer way, namely through a penalty term. Our results and analysis are analogous.

Finally using the insights of our analysis, we define a modified model which is similar to the original one, but for which the asymptotic consistency holds as $n \to \infty$ with no other upper bound on ε_n other than $\varepsilon_n \to 0$ as $n \to \infty$.

We remark that while we consider measures on open sets in $\mathbb{R}^d$, there are no essential difficulties to extend the results to manifold setting, namely one where μ is a measure supported on compact manifold $\mathcal{M}$ of dimension d embedded in $\mathbb{R}^D$. This is already being done for the Laplacian [27] where the modification of background results (such as optimal transportation estimates) has been carried out.

To prove our results we use the tools of calculus of variations and optimal transportation. In particular we use the setup for convergence of objective functionals defined on graphs to their continuum limits developed in [30]. This includes the definition of the proper topology (TL^p) to compare functionals defined on finite discrete objects (graphs) with their continuum limits. However the TL^p topology, which is an extension of the L^p topology, is not strong enough to ensure that the labels are preserved in the limit. For this reason we also need to consider a stronger topology, namely the one of uniform convergence. Proving the needed local regularity results for the discrete p-Laplacian (Lemma 4.1) and the compactness results needed to ensure the locally uniform convergence are the main technical contributions of the paper.

The paper is organized as follows. We complete the introduction with a review on related works. In Section 2 we give a precise description of the problem with the assumptions and state the main

results. Section 3 contains a brief overview of background results we use. This includes a description of the TL^p topology, which we use for discrete-to-continuum convergence, and a short overview of Γ-convergence and optimal transportation. Section 4 contains the proofs of the main results given in Section 2. In Section 5 we present an improved model that, while similar to the constrained problem for $\mathcal{E}_n^{(p)}(f)$, is asymptotically consistent with the desired limiting problem even when $\varepsilon_n \to 0$ slowly as $n \to \infty$ We conclude the paper with 1D numerical examples in Section 6.

1.1 Discussion of Related Works

The approach to semi-supervised learning using a weighted graph to represent the geometry of the unlabeled data and Laplacian based regularization was proposed by Zhu, Ghahramani, and Lafferty in [59]. It fits in the general theme of graph-Laplacian based approaches to machine learning tasks such as clustering, which are reviewed in [54]. Zhou and Schölkopf [57] generalized the regularizers of [59] to include a version of the graph p-Laplacian. The p-Laplacian regularization has also been used by Bühler and Hein in clustering problems [8], where values of p close to 1 are of particular interest due to connections with graph cuts. Graph based p-Laplacian regularization has found further applications in semi-supervised learning and image processing [15–17]. These papers also make the connection to the ∞-Laplacian, which is closely related to minimal Lipschitz extensions [11].

While the approach of [59] has found many applications it was pointed out by Nadler, Srebro and Zhou [39] that the estimator degenerates and becomes uninformative in $d \geq 2$, when the number of unlabeled data points $n \to \infty$. Almagir and von Luxburg [2] explored the p-resistances, the resulting distance on graphs, and connections to the p-Laplace regularization. Based on their analysis they suggested that $p = d$ should be a good choice to prevent degeneracy in the $n \to \infty$ limit. El Alaoui, Cheng, Ramdas, Wainwright, and Jordan [14] show that for $p \leq d$ the problem degenerates as $n \to \infty$ and spikes can occur. They argue that regularizations with high $p \geq d + 1$ are sufficient to prevent the appearance of spikes as $n \to \infty$, and lead to a well-posed problem in the limit. Here we make part of their claims rigorous, namely that if $p > d$ then the asymptotic consistency holds only if ε_n converges to zero sufficiently fast ($\varepsilon_n n^p \to 0$ as $n \to \infty$). If $p > d$ and $\varepsilon_n n^p \to \infty$ as $n \to \infty$ we prove that the problem still degenerates as $n \to \infty$ and that spikes occur. We also introduce a modification to the discrete problem (by modifying how the agreement with the assigned labels is imposed) which is well posed when $p > d$ without the need for ε_n to converge to 0 quickly.

There are other ways to regularize the SSL regression problems which ensure that no spikes occur. Namely Belkin and Niyogi [4,5] consider estimators which are required to lie in the space spanned by a fixed number of eigenvectors of the graph Laplacian. Due to the smoothness of low eigenvectors of the Laplacian this prevents the formation of spikes. One can think of this approach in energy based setting where infinite penalty has been imposed on high frequencies. A softer, but still linear, way to do this is to consider (fractional) powers of the graph Laplacian, namely the regularity term $J_n(u) = \langle cL_n^\alpha f, f \rangle$ where L_n is the graph Laplacian, and $\alpha > 0$. This regularization was studied by Belkin and Zhou [58] who argue, again via regularity obtained by Sobolev embedding theorems, that taking $\alpha > \frac{d}{2}$ prevents spikes. However Dunlop, Stuart, and the authors have discovered that a similar phenomenon to one described in this paper. Namely even when $\alpha > \frac{d}{2}$ the limit may be degenerate, and spikes can occur, if ε_n converges to zero slowly, namely if $\varepsilon_n^{2\alpha} n \to \infty$ as $n \to \infty$.

Our results fall in the class of asymptotic consistency results in machine learning. In general one is interested the asymptotic behavior of an objective function posed on a random sample of n points, and which also depends on a parameter ε, $E_{n,\varepsilon}(f_n)$ where f_n is a real valued function defined at sample

points. The limit is considered as $n \to \infty$ while $\varepsilon_n \to 0$ at appropriate rate. The limiting problem is described by a continuum functional $E_\infty(f)$ which acts on real valued functions supported on domains or manifolds. Also relevant is the (nonlocal) continuum problem, $E_{\infty,\varepsilon}(f)$ which describes the limit $n \to \infty$ while $\varepsilon > 0$ is kept fixed.

The type of consistency that is needed for the conclusions, and the one we consider, is *variational consistency*, namely that minimizers of $E_{n,\varepsilon_n}(f_n)$ converge to minimizers of $E_\infty(f)$ as $n \to \infty$ while $\varepsilon_n \to 0$ at an appropriate rate. Proving such results includes choosing the right topology to compare the functions on discrete domain f_n with those on the continuum domain f.

Many works in the literature are interested in a simpler notion of convergence, namely that for a fixed, sufficiently smooth, continuum function f it holds that $E_{n,\varepsilon_n}(f) \to E_\infty(f)$ as $n \to \infty$ while $\varepsilon_n \to 0$ at an appropriate rate, where by $E_{n,\varepsilon_n}(f)$ we mean that the discrete functional is evaluated at the restriction of f to the data points. We call this notion of convergence *pointwise convergence*. A somewhat weaker notion of convergence is what we here call *iterated pointwise convergence*, namely considering $\lim_{\varepsilon\to 0} \lim_{n\to\infty} E_{n,\varepsilon}(f)$. Also relevant for the problems based on linear operators (namely the graph Laplacian) is *spectral convergence* which asks for the eigenvalues and eigenvectors of the discrete operator to converge to eigenvalues and eigenfunction of the continuum one. This notion of the convergence is typically sufficient for the kind of conclusions we are investigating (however our problems are nonlinear).

Pointwise (and similar notions of) convergence of graph Laplacians was studied by Belkin and Niyogi [6], Coifman and Lafon [10], Giné and Koltchinskii [33], Hein, Audibert and von Luxburg [35], Hein [34], Singer [45], and Ting, Huang, and Jordan [52]. Spectral convergence was studied in the works of Belkin and Niyogi [6] on the convergence of Laplacian eigenmaps, von Luxburg, Belkin, and Bousquet [55] and Pelletier and Pudlo [40] on graph Laplacians, and of Singer and Wu [46] on the connection graph Laplacian. In these works on spectral convergence either ε remains fixed as $n \to \infty$ or $\varepsilon(n) \to 0$ at an unspecified rate. The precise and almost optimal rates were obtained in [31] using variational methods. Further problems involve obtaining error estimates between discrete and continuum objects. Laplacians on discretized manifolds was studied by Burago, Ivanov and Kurylev [9] who obtain precise error estimates for eigenvalues and eigenvectors. Related results on approximating elliptic equations on point clouds have been obtained by Li and Shi [37], and Li, Shi, Sun, [38]. Error bounds for the spectral convergence of graph Laplacians have been considered by Wang [56] and García-Trillos, Gerlach, Hein and one of the authors [27]. Regarding graph p-Laplacians, which are the subject of this work, the authors of [14] obtain iterated pointwise convergence of graph p-Laplacians to the continuum p-Laplacian.

To obtain the results on variational convergence of $\mathcal{E}_n^{(p)}$ to $\mathcal{E}_\infty^{(p)}$ needed to fully explain the asymptotics of discrete regression problems we combine tools of calculus of variations (in particular Γ-convergence) and optimal transportation. This approach to asymptotics of problems posed on discrete random samples was developed by García-Trillos and one of the authors [30,31]. In [30] they introduce the TL^p topology for comparing the functions defined on the discrete sets to the ones defined in the continuum, and apply the approach to asymptotics of graph-cut based objective functions. We refer to this paper for a description of the rich background of the works that underpin the approach. In [31] the authors apply the approach to convergence of graph Laplacian based functionals. Consistency of k-means clustering for paths with regularization was recently studied by Theil, Johansen and Cade, and one of the authors [51], using a similar viewpoint. This technical setup has recently been used and extended to studies on modularity based clustering [13], Cheeger and ratio cuts [32], neighborhood graph constructions for graph cut based clustering [26], and classification problems [28, 50].

An alternative approach to related regression problems was developed by Fefferman and collaborators Israel, Klartag and Luli, who look for a function of sufficient regularity, e.g. C^m or $W^{m,p}$, that extends a function $f^\dagger : E \to \mathbb{R}$ to the whole of $\mathbb{R}^d$ in such a way as to minimize the norm of the extension. Considerable work has gone into showing such extensions exist and finding constructions for f, e.g. for C^m regularity [19,23,24], and for Sobolev regularity [20–22]. In the context of machine learning this is a supervised learning problem and thus only makes use of the labeled data. In our context the problem is independent of $\{x_i\}_{i=N+1}^n$ and does not make use of the geometry given by the unlabeled data.

2 Setting and Main Results

Let Ω be an open, bounded domain in $\mathbb{R}^d$. Let $N \geq 0$ and let $\{(x_i, y_i) \,:\, i = 1, \ldots, N\}$ with $x_i \in \Omega$ and $y_i \in \mathbb{R}$ be a collection of distinct labeled points. We consider μ to be the measure representing the distribution of data. We assume that $\mathrm{supp}\mu = \overline{\Omega}$ and that μ has density ρ with respect to Lebesgue measure. We assume that ρ is continuous and is bounded above and below by positive constants on Ω.

We assume that unlabeled data, $\{x_i\}_{i=N+1,\ldots}$ are given by a sequence of iid samples of measure μ. The empirical measure induced by data points is given by $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. Let $G_n = (\Omega_n, E_n, W_n)$ be a graph with vertices $\Omega_n = \{x_i \,:\, i = 1, \ldots, n\}$, edges $E_n = \{e_{ij}\}_{i,j=1}^n$ and edge weights $W_n = \{W_{ij}\}_{i,j=1}^n$. For notational simplicity we will set $W_{ij} = 0$ if there is no edge between x_i and x_j.

We assume the following structure on edge weights

$$W_{ij} = \eta_\varepsilon(|x_i - x_j|) \tag{1}$$

where $\eta_\varepsilon(|x|) = \frac{1}{\varepsilon^d}\eta\left(\frac{|x|}{\varepsilon}\right)$, $\eta : [0,\infty) \to [0,\infty)$ is a nonincreasing kernel and $\varepsilon = \varepsilon_n$ is a scaling parameter depending on n. For example if $\eta(|x|) = \mathbb{I}_{|x|\leq 1}$ then $\eta_\varepsilon(|x|)$ is $\frac{1}{\varepsilon^d}$ if $|x| \leq \varepsilon$ and 0 otherwise. In this case vertices are only connected if they are closer than ε.

We consider two models: one where the agreement of the response with the training variables is imposed as a constraint and the other where it is imposed via quadratic penalty. We call these models *constrained* and *penalized* respectively.

In the constrained model we construct our estimator as the minimizer of

$$\mathcal{E}_n^{(p)}(f) = \frac{1}{\varepsilon_n^p}\frac{1}{n^2}\sum_{i,j=1}^n W_{ij}|f(x_i) - f(x_j)|^p \tag{2}$$

among $\{f : \Omega_n \to \mathbb{R}\}$ which satisfy the constraint $f(x_i) = y_i$ for all $i = 1, \ldots, N$.

For technical reasons it is more convenient to define the functional over all f and impose the constraint in the following way

$$\mathcal{E}_{n,con}^{(p)}(f) = \begin{cases} \frac{1}{\varepsilon_n^p}\frac{1}{n^2}\sum_{i,j=1}^n W_{ij}|f(x_i) - f(x_j)|^p & \text{if } f(x_i) = y_i \text{ for } i = 1, 2, \ldots, N \\ \infty & \text{else.} \end{cases} \tag{3}$$

We now turn to the penalized formulation. For $q > 0$ let

$$R^{(q)}(f) = \sum_{i=1}^N |y_i - f(x_i)|^q.$$

We define the estimator as the minimizer of

$$\mathcal{S}_n^{(p)}(f) = \mathcal{E}_n^{(p)}(f) + \lambda R^{(q)}(f) \tag{4}$$

where $\lambda > 0$ is a tunable parameter.

We now introduce the continuum functionals that describe the limiting problems as $n \to \infty$. Let

$$\mathcal{E}_\infty^{(p)}(f) = \begin{cases} \sigma_\eta \int_\Omega |\nabla f(x)|^p \, \rho^2(x) \, \mathrm{d}x & \text{if } f \in W^{1,p}(\Omega), \\ \infty & \text{else.} \end{cases} \tag{5}$$

and for $p > d$, Sobolev functions $f \in W^{1,p}$ are continuous and we can define

$$\mathcal{E}_{\infty,con}^{(p)}(f) = \begin{cases} \sigma_\eta \int_\Omega |\nabla f(x)|^p \, \rho^2(x) \, \mathrm{d}x & \text{if } f \in W^{1,p}(\Omega) \text{ and } f(x_i) = y_i \text{ for } i = 1, \dots, N \\ \infty & \text{else.} \end{cases} \tag{6}$$

To describe the limit of the penalized model the large data limit we introduce

$$\mathcal{S}_\infty^{(p)}(f) = \mathcal{E}_\infty^{(p)}(f) + \lambda R^{(q)}(f) \tag{7}$$

In the above σ_η is the constant

$$\sigma_\eta = \int_{\mathbb{R}^d} \eta(|x|) \, |x \cdot e_1|^p \, \mathrm{d}x,$$

where $e_1 = [1, 0, \dots, 0]^T$.

We note that both functionals (6) and (7) are lower semi-continuous with respect to the L^p metric. In addition, coercivity of both functionals follows from Sobolev embeddings. Corecivity and lower semi-continuity imply existence of minimizers, e.g. [25, Theorem 3.6]. Furthermore, strict convexity implies the minimizers are unique.

We are interested in asymptotic behavior of minimizers f_n of the discrete models, say $\mathcal{E}_{n,con}^{(p)}$. We say that $\mathcal{E}_{n,con}^{(p)}$ is *asymptotically consistent* with $\mathcal{E}_{\infty,con}^{(p)}$ if the minimizers f_n of $\mathcal{E}_{n,con}^{(p)}$ converge as $n \to \infty$ to a minimizer of $\mathcal{E}_{\infty,con}^{(p)}$. One should note topology of the convergence $f_n \to f_\infty$ is not at this stage clear.

We observe that since $f_n : \Omega_n \to \mathbb{R}$, while $f : \Omega \to \mathbb{R}$ this issue is nontrivial. We use the TL^p topology introduced in [30] precisely to compare functions defined on different domains in a topology consistent with L^p convergence. We define the convergence rigorously in Section 3.

Another issue is the rate at which ε_n is allowed to converge to zero. If $\varepsilon_n \to 0$ too quickly then the graph becomes disconnected and hence it does not capture the geometry of Ω properly. The connectivity threshold [41] is $\varepsilon_n \sim \left(\frac{\log n}{n}\right)^{\frac{1}{d}}$. We require (when $d \geq 3$) $\varepsilon_n \gg \left(\frac{\log n}{n}\right)^{\frac{1}{d}}$ which means that our lower bound needed is almost optimal. On the other hand we discovered that if $\varepsilon_n \to 0$ too slowly the discrete functional $\mathcal{E}_{n,con}^{(p)}$ lacks sufficient regularity for the constraints to be preserved in the limit. The optimal upper bound on ε_n is discussed in Theorem 2.1.

We now state our assumptions needed for the main results.

(A1) $\Omega \subset \mathbb{R}^d$ is open, connected, bounded and with Lipschitz boundary;

(A2) The probability measure $\mu \in \mathcal{P}(\Omega)$ has continuous density ρ which is bounded above and below by strictly positive constants in Ω;

(A3) There exists N labeled points: $(x_i, y_i) \in \Omega \times \mathbb{R}$ for $i = 1, \dots, N$;

(A4) For $i > N$ the data points x_i, are iid samples of μ;

(A5) Let ε_n be a sequence converging to 0 satisfying the lower bound

$$\varepsilon_n \gg \begin{cases} \sqrt{\frac{\log\log n}{n}} & \text{if } d = 1 \\ \frac{(\log n)^{\frac{3}{4}}}{\sqrt{n}} & \text{if } d = 2 \\ \left(\frac{\log n}{n}\right)^{\frac{1}{d}} & \text{if } d \geq 3; \end{cases}$$

(A6) The kernel profile $\eta : [0, \infty) \to [0, \infty)$ is non-increasing;

(A7) η is positive and continuous at $x = 0$;

(A8) The integral $\int_0^\infty \eta(t)|t|^{p+d} \, \mathrm{d}t$ is finite (equivalently $\sigma_\eta = \int_{\mathbb{R}^d} \eta(|w|)|w \cdot e_1|^p \, \mathrm{d}w < \infty$).

The first main result of the paper is the following theorem. Its proof is presented in Section 4.

Theorem 2.1 (Consistency of the constrained model)**.** *Let $p > 1$. Assume Ω, μ, η, and x_i satisfy the assumptions **(A1) - (A8)**. Let graph weights W_{ij} be given by* (1)*. Let f_n be a sequence of minimizers of $\mathcal{E}^{(p)}_{n,con}$ defined in* (6)*. Then, almost surely, the sequence (μ_n, f_n) is precompact in the TL^p metric. The TL^p limit of any convergent subsequence, (μ_{n_m}, f_{n_m}), is of the form (μ, f) where $f \in W^{1,p}(\Omega)$. Furthermore,*

(i) if $n\varepsilon_n^p \to 0$ as $n \to \infty$ then f is continuous and

(a) f_{n_m} converges locally uniformly to f, meaning that for any $\Omega' \subset\subset \Omega$

$$\lim_{m\to\infty} \max_{\{k \leq n_m \,:\, x_k \in \Omega'\}} |f(x_k) - f_{n_m}(x_k)| = 0,$$

(b) f is a minimizer of $\mathcal{E}^{(p)}_{\infty,con}$ defined in (6)*,*

(c) the whole sequence f_n converges to f both in TL^p and locally uniformly;

(ii) if $n\varepsilon_n^p \to \infty$ as $n \to \infty$ then f is a minimizer of $\mathcal{E}^{(p)}_{\infty}$ defined in (5)*.*

We note that in case (i) assumption **(A5)** and $n\varepsilon_n^p \to 0$ as $n \to \infty$ imply that $n^{-1/p} \gg \varepsilon \gg n^{-1/d}$ which is only possible if $p > d$. Therefore in case (i) we always have that functions f for which $\mathcal{E}^{(p)}_{\infty}$ is finite are always continuous and thus it is possible to impose pointwise values of f, as needed to define $\mathcal{E}^{(p)}_{\infty,con}$ in (6).

The result (i) establishes the asymptotic consistency of the discrete constrained model with the constrained continuum weighted p-Laplacian model.

While the result (ii) looks similar its interpretation is different. It shows that the model "forgets" the constraints in the limit. Namely $\mathcal{E}^{(p)}_{\infty}$ only has the gradient term and no constraints! In particular its minimizers are constants over Ω. What is happening is that f_n develops narrow spikes near the labeled points x_i and becomes nearly constant everywhere else. In the TL^p limit the spikes disappear.

This motivates referring to the scaling when $n^p\varepsilon \to \infty$ as $n \to \infty$ as the ***degenerate*** regime. On the other hand, we refer to the scaling of case (i) as the ***well-posed*** regime.

The other main result is the convergence in the penalized model. The proof is a straightforward extension of Theorem 2.1 in the special case $N = 0$ (so that the constraint is not present). We include the proof in Section 4.2.

Proposition 2.2. *Let $p > 1$. Assume Ω, μ, η, and x_i satisfy the assumptions **(A1)-(A8)**. Let graph weights W_{ij} be given by* (1)*. Let f_n be a sequence of minimizers of $\mathcal{S}_n^{(p)}$ defined in* (4)*. Then, almost surely, the sequence (μ_n, f_n) is precompact in the TL^p metric. The TL^p limit of any convergent subsequence, (μ_{n_m}, f_{n_m}), is of the form (μ, f) where $f \in W^{1,p}(\Omega)$. Furthermore,*

(i) if $n\varepsilon_n^p \to 0$ as $n \to \infty$ then f is continuous and

(a) f_n converges locally uniformly to f, meaning that for any $\Omega' \subset\subset \Omega$

$$\lim_{n\to\infty} \max_{\{k \le n \,:\, x_k \in \Omega'\}} |f(x_k) - f_n(x_k)| = 0,$$

(b) f is a minimizer of $\mathcal{S}_\infty^{(p)}$ defined in (6)*,*

(c) the whole sequence f_n converges to f both in TL^p and locally uniformly;

(ii) if $n\varepsilon_n^p \to \infty$ as $n \to \infty$ then f is a minimizer of $\mathcal{E}_\infty^{(p)}$ defined in (5)*.*

Again the result of (i) is a consistency result, while (ii) shows that the penalization of the labels is lost in the limit.

Remark 2.3. The above results (Theorem 2.1 and Proposition 2.2) could also be extended to $p = 1$, in which case the limiting functional $\mathcal{E}_\infty^{(1)}$ would be a weighted TV semi-norm $\mathcal{E}_\infty^{(1)} = \sigma_\eta TV(\cdot; \rho)$ where

$$TV(f; \rho) = \sup \left\{ \int_\Omega f \mathrm{div}\phi \, \mathrm{d}x \,:\, |\phi(x)| \le \rho^2(x) \, \forall x \in \Omega, \phi \in C_c^\infty(\Omega; \mathbb{R}^d) \right\}.$$

A modification of the proofs contained here would prove the result, see also [30].

We conclude this section with a short discussion on the heuristics behind each of the three scenarios: (a) $p \le d$, (b) $p > d$ and $n\varepsilon_n^p \to \infty$, and (c) $p > d$ and $n\varepsilon_n^p \to 0$. We note that (a) and (b) are in the degenerate regime, while (c) is in the well-posed regime.

When $p \le d$ there is no embedding of $W^{1,p}$ into the continuous functions and moreover one cannot define a trace of a $W^{1,p}$ function in $\mathbb{R}^d$ with $d > 1$ at a point. Since the functional $\mathcal{E}_\infty^{(p)}$ provides no further control than the $W^{1,p}$ norm one cannot expect to be able to impose pointwise data in this case. So the failure of consistency is due to the lack of regularity of the limiting problem and is not surprising.

When $p > d$ but ε_n decreases slowly then the averaging effect of convolving the finite differences with η_{ε_n}, which averages over a length scale ε_n, means that even though the convolved finite differences are smooth this does not pass to the underlying functions. Essentially this allows spikes to grow on a scale smaller than ε_n without paying a price in $\mathcal{E}_n^{(p)}$. Even though (by Morrey's inequality) functions f with $\mathcal{E}_\infty(f) < \infty$ are continuous and therefore pointwise evaluation is well defined, the constraints do not survive in the limit.

In the final case, when $p > d$ and ε_n decreases sufficiently quickly, we are able to show that the constraints are satisfied. The constraints clearly make sense (since pointwise evaluation is well defined whenever $p > d$), and furthermore ε_n decreases sufficiently quickly to ensure that spikes pay a high price in $\mathcal{E}_n^{(p)}$.

We propose an improved model in Section 5 that is well-posed for all $p > d$.

3 Background Material

In an effort to make this paper more self-contained we briefly recall three key notions our work relies on. The first is Γ-convergence which is a notion of convergence of functionals developed for the analysis of sequences of variational problems. The second is the notion of optimal transportation, and the third is the TL^p space which we use to define the convergence of discrete functions to continuum functions.

3.1 Γ–Convergence

Γ-convergence was introduced by De Giorgi in 1970's to study limits of variational problems. We refer to [7, 12] for an in depth introduction to Γ-convergence. Our application of Γ-convergence will be in a random setting.

Definition 3.1 (Γ-convergence)**.** *Let (Z, d) be a metric space and $(\mathcal{X}, \mathbb{P})$ be a probability space. For each $\omega \in \mathcal{X}$ the functional $E_n^{(\omega)} : Z \to \mathbb{R} \cup \{\pm\infty\}$ is a random variable. We say $E_n^{(\omega)}$ Γ-converge almost surely on the domain Z to $E_\infty : Z \to \mathbb{R} \cup \{\pm\infty\}$ with respect to d, and write $E_\infty = \Gamma\text{-}\lim_{n\to\infty} E_n^{(\omega)}$, 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 $\{f_n\}_{n=1}^\infty$ converging to f

$$E_\infty(f) \leq \liminf_{n\to\infty} E_n^{(\omega)}(f_n), \text{ and}$$

(ii) (recovery sequence) there exists a sequence $\{f_n\}_{n=1}^\infty$ converging to f such that

$$E_\infty(f) \geq \limsup_{n\to\infty} E_n^{(\omega)}(f_n).$$

For ease of notation we will suppress the dependence of ω on on our functionals, that is we apply the above definition to $E_n = \mathcal{E}_n^{(p)}$. The almost sure statement in the above definition does not play a significant role in the proofs. Basically it is enough to consider the set of realisations of $\{x_i\}_{i=1}^\infty$ such that the empirical measure converges weak*. More precisely, we consider the set of realizations of $\{x_i\}_{i=1}^\infty$ such that the conclusions of Theorem 3.3 hold.

The fundamental result concerning Γ-convergence is the following convergence of minimizers result. The proof can be found in [7, Theorem 1.21] or [12, Theorem 7.23].

Theorem 3.2 (Convergence of Minimizers)**.** *Let (Z, d) be a metric space and $E_n : Z \to [0, \infty]$ be a sequence of functionals. Let f_n be a minimizing sequence for E_n. If the set $\{f_n\}_{n=1}^\infty$ is precompact and $E_\infty = \Gamma\text{-}\lim_n E_n$ where $E_\infty : Z \to [0, \infty]$ is not identically $+\infty$ then*

$$\min_Z E_\infty = \lim_{n\to\infty} \inf_Z E_n.$$

Furthermore any cluster point of $\{f_n\}_{n=1}^\infty$ is a minimizer of E_∞.

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

We note that Γ-convergence is defined for functionals on a common metric space. The next section 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.

3.2 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 [53] and Santambrogio [43].

Given Ω is open and bounded, and probability measures μ and ν in $\mathcal{P}(\overline{\Omega})$ we define the set $\Pi(\mu,\nu)$ of transportation maps, or couplings, between them 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

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

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

$$d_p(\mu,\nu) = \begin{cases} \inf\limits_{\pi\in\Pi(\mu,\nu)} \left(\int_{\Omega} |x-T(x)|^p \,\mathrm{d}\mu(x)\right)^{\frac{1}{p}} & \text{if } 1\le p<\infty \\ \inf\limits_{T_\sharp\mu=\nu} \mu\text{-}\operatorname*{ess\,sup}\limits_{x} |x-T(x)| & \text{if } p=\infty. \end{cases}$$

where $T_\mu=\nu$ means that the push forward of measure μ by T is 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 $d\ge 2$, they were established in [29], which extended the related results in [1, 36, 44, 47]. For $d=1$ the estimates are simpler, and follow from the law of iterated logarithms.

Theorem 3.3. *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 $x_1,x_2,\dots$ be a sequence of independent random variables with distribution μ and let μ_n be the empirical measure. Then there exists a constants $C\ge c>0$ such that almost surely there exists a sequence of transportation maps $\{T_n\}_{n=1}^\infty$ from μ to μ_n such that*

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

where

$$\delta_n = \begin{cases} \sqrt{\frac{\log\log(n)}{n}} & \text{if } d=1 \\ \frac{(\log n)^{\frac{3}{4}}}{\sqrt{n}} & \text{if } d=2 \\ \frac{(\log n)^{\frac{1}{d}}}{n^{\frac{1}{d}}} & \text{if } d\ge 3. \end{cases}$$

3.3 The TL^p Space

The discrete functionals we consider (e.g. $\mathcal{E}_n^{(p)}$) are defined for functions $f_n : \Omega_n \to \mathbb{R}$ where $\Omega_n = \{x_i \ : \ i = 1, \dots, n\}$, while the limit functional $\mathcal{E}_\infty^{(p)}$ acts on functions $f : \Omega \to \mathbb{R}$, where Ω is an open set. We can view f_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 out of which the points are sampled from. One would like how to compare f and f_n in a way that is consistent with L^p topology. To do so we use the TL^p space was introduced in [30], where it was used to study the continuum limit of the graph total variation (that is $\mathcal{E}_n^{(1)}$). Subsequent development of the TL^p space has been carried out in [31, 48, 49].

To compare the functions f_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

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

where $\Pi(\mu, \nu)$ the set of transportation plans defined in Section 3.2. We note that the minimizing π exists and that TL^p space is a metric space, [30].

When μ has a density with respect to Lebesgue measure on Ω, then the distance can be rewritten using transportation maps, T instead of transportation plans,

$$d_{TL^p}^p((\mu, f), (\nu, g)) = \inf\left\{\int_{\Omega} |x - T(x)|^p + |f(x) - g(T(x))|^p \, \mathrm{d}\mu(x) \ : \ T_\# \mu = \nu\right\}.$$

This formula provides a clear interpretation of the distance in our setting. Namely to compare functions $f_n : \Omega_n \to \mathbb{R}$ we define a mapping $T_n : \Omega \to \Omega_n$ and compare the functions $\tilde{f}_n = f_n \circ T_n$ and f in $L^p(\mu)$, while also accounting for the transport, namely the $|x - T_n(x)|^p$ term.

We remark that $TL^p(\overline{\Omega})$ space is not complete and that its completion was discussed in [30]. In the setting of this paper, since the corresponding measure is clear from context, we often say that f_n converges in TL^p to f as a short way to say that (μ_n, f_n) converges in TL^p to (μ, f).

4 Regularity and asymptotics of discrete and nonlocal functionals

Here we present some of the key properties of the functionals involved that allow us to show the asymptotic consistency of Theorem 2.1. A fundamental new issue (compared to say [31]) is that constraints in $\mathcal{E}_\infty^{(p)}$ are imposed pointwise on a set of μ measure zero. [The reason that these constraints make sense is that for $p > d$ finiteness of $\mathcal{E}_\infty^{(p)}(f)$ implies that f is continuous.] We note that the TL^p convergence used in [31] is not sufficient to imply that constraints are preserved. One needs a stronger convergence, like the uniform one. This raises the question on how to obtain the needed compactness of sequences f_n for which $\mathcal{E}_{n,con}^{(p)}(f_n)$ is uniformly bounded. Our approach combines discrete and continuum regularity results. Namely we obtain in Lemma 4.1 a local control of oscillation of f_n over distances of order ε_n. In Lemma 4.2 we show that discrete functionals $\mathcal{E}_{n,con}^{(p)}(f_n)$ control the values of the associated nonlocal continuum functionals $\mathcal{E}_{\varepsilon_n}^{(NL,p)}(\tilde{f}_n)$ (defined in (10) below) applied to an

appropriate extrapolation $\tilde{f}_n$ of f_n. A simple but important point is that the discrete functionals at fixed n are always closer to a nonlocal functional with nonlocality at scale ε_n, than to the limiting functional. The issue is that these nonlocal functionals do not share the regularizing properties of the limiting functional. However we show in Lemma 4.3 that control of the nonlocal energy is sufficient to provide regularity at scales larger than ε. Combining these estimates is enough to imply the compactness with respect to (locally) uniform convergence, Lemma 4.5.

Lemma 4.1 (Discrete regularity)**.** *Let $p > 1$. Assume Ω, μ, η, and x_i satisfy the assumptions **(A1) - (A8)**. Let graph weights W_{ij} be given by* (1)*. Let $\Omega_n = \{x_i\}_{i=1}^n$. For $f_n : \Omega_n \to \mathbb{R}$, define $\operatorname{osc}_\varepsilon^{(n)}(f_n) : \Omega_n \to \mathbb{R}$ by*

$$\operatorname{osc}_\varepsilon^{(n)}(f_n)(x_i) = \max_{z \in B(x_i,\varepsilon)\cap\Omega_n} f_n(z) - \min_{z \in B(x_i,\varepsilon)\cap\Omega_n} f_n(z).$$

For any $\alpha_0 > 0$, with probability one, there exist $n_0 > 0$ and $C > 0$ (independent of n) such that for any $\alpha \geq \alpha_0$, all $n \geq n_0$ and $k \in \{1, 2, \ldots, n\}$:

$$\left(\operatorname{osc}_{\alpha\varepsilon_n}^{(n)}(f_n)(x_k)\right)^p \leq C\alpha^p n\varepsilon_n^p \mathcal{E}_n^{(p)}(f_n),$$

where $\mathcal{E}_n^{(p)}$ is defined by (2)

Proof. Let $\tilde{\eta}(t) = a$ if $0 \leq t < b$ and $\tilde{\eta}(t) = 0$ where a and b are chosen such that $\tilde{\eta} \leq \eta$. We can furthermore choose b so that $b \leq \alpha_0$. For all $k \in \{1, \ldots, n\}$ let

$$\bar{f}_n(x_k) = \max_{z \in B(x_k, \frac{b\varepsilon_n}{2})\cap\Omega_n} f_n(z), \qquad \bar{x}_k \in \operatorname*{argmax}_{z \in B(x_k, \frac{b\varepsilon_n}{2})\cap\Omega_n} f_n(z),$$

$$\underline{f}_n(x_k) = \min_{z \in B(x_k, \frac{b\varepsilon_n}{2})\cap\Omega_n} f_n(z), \qquad \underline{x}_k \in \operatorname*{argmin}_{z \in B(x_k, \frac{b\varepsilon_n}{2})\cap\Omega_n} f_n(z).$$

Note that $\operatorname{osc}_{\frac{b\varepsilon_n}{2}}^{(n)}(f_n)(x_k) = \bar{f}_n(x_k) - \underline{f}_n(x_k)$ and for all $x \in B\left(x_k, \frac{b\varepsilon_n}{2}\right) \cap \Omega_n$

$$\text{(i)} \qquad f_n(x) - \bar{f}_n(x_k) \geq \frac{1}{2}\operatorname{osc}_{\frac{b\varepsilon_n}{2}}(f_n)(x_k),$$

$$\text{or} \quad \text{(ii)} \qquad f_n(x) - \underline{f}_n(x_k) \geq \frac{1}{2}\operatorname{osc}_{\frac{b\varepsilon_n}{2}}(f_n)(x_k).$$

Without a loss of generality we assume that (i) holds for at least half the points in $B\left(x_k, \frac{b\varepsilon_n}{2}\right) \cap \Omega_n$. Then,

$$\begin{aligned}
\mathcal{E}_n^{(p)}(f_n) &\geq \frac{1}{\varepsilon_n^{p+d} n^2} \sum_{i,j=1}^n \tilde{\eta}\left(\frac{|x_i - x_j|}{\varepsilon_n}\right) |f_n(x_i) - f_n(x_j)|^p \\
&\geq \frac{1}{\varepsilon_n^{p+d} n^2} \sum_{j:|x_j - \bar{x}_k| \leq b\varepsilon_n} \tilde{\eta}\left(\frac{|\bar{x}_k - x_j|}{\varepsilon_n}\right) |f_n(x_j) - f_n(\bar{x}_k)|^p \\
&\geq \frac{a}{\varepsilon_n^{p+d} n^2} \sum_{j:|x_j - x_k| \leq \frac{b\varepsilon_n}{2}} |f_n(x_j) - f_n(\bar{x}_k)|^p, \qquad \text{since } |x_k - \bar{x}_k| \leq \frac{b\varepsilon_n}{2} \qquad (8) \\
&\geq \frac{a}{2^{p+1}\varepsilon_n^{p+d} n^2} \left(\operatorname{osc}_{\frac{b\varepsilon_n}{2}}(f_n)(x_k)\right)^p \# \left\{j \,:\, |x_j - x_k| \leq \frac{b\varepsilon_n}{2}\right\} \\
&= \frac{a}{2^{p+1}\varepsilon_n^{p+d} n} \left(\operatorname{osc}_{\frac{b\varepsilon_n}{2}}(f_n)(x_k)\right)^p \mu_n\left(B\left(x_k, \frac{b\varepsilon_n}{2}\right)\right).
\end{aligned}$$

where $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. Now, for a transport map $T_n : \Omega \to \Omega_n$ from μ to μ_n, satisfying the conclusions of Theorem 3.3, we have

$$\begin{aligned}\frac{1}{\varepsilon_n^d}\mu_n\left(B\left(x_k, \frac{b\varepsilon_n}{2}\right)\right) &= \frac{1}{\varepsilon_n^d}\int_\Omega \mathbb{I}_{\{|T_n(x)-x_k|\le \frac{b\varepsilon_n}{2}\}}\rho(x)\,\mathrm{d}x \\ &\ge \frac{\inf_{x\in\Omega}\rho}{\varepsilon_n^d}\int_\Omega \mathbb{I}_{\{|x-x_k|\le \frac{b\varepsilon_n}{2}-\|T_n-\mathrm{Id}\|_{L^\infty}\}}\,\mathrm{d}x \\ &= \big(\inf_{x\in\Omega}\rho(x)\big)\mathrm{Vol}\left(B\left(0, \frac{b}{2}-\frac{\|T_n-\mathrm{Id}\|_{L^\infty}}{\varepsilon_n}\right)\right).\end{aligned} \tag{9}$$

We choose n_0 such that for $n \ge n_0$ it holds that $\frac{\|T_n-\mathrm{Id}\|_{L^\infty}}{\varepsilon_n} \le \frac{b}{4}$. Combining (8) and (9) gives

$$\left(\mathrm{osc}_{\frac{b\varepsilon_n}{2}}(f_n)(x_k)\right)^p \le \frac{2^{p+1}\varepsilon_n^p n\mathcal{E}_n^{(p)}(f_n)}{a\big(\inf_{x\in\Omega}\rho(x)\big)\mathrm{Vol}\left(B\left(0,\frac{b}{4}\right)\right)} =: C_1\varepsilon_n^p n\mathcal{E}_n^{(p)}(f_n).$$

For $\alpha > \alpha_0$, using $\alpha_0 \ge b$ and applying the triangle inequality $\left\lfloor \frac{2\alpha}{b} \right\rfloor$ times, we obtain

$$(\mathrm{osc}_{\alpha\varepsilon_n}(f_n)(x_k))^p \le C_1\left(\left\lfloor\frac{2\alpha}{b}\right\rfloor + 1\right)^p \varepsilon_n^p n\mathcal{E}_n^{(p)}(f_n) \le C_1\left(\frac{3\alpha}{b}\right)^p \varepsilon_n^p n\mathcal{E}_n^{(p)}(f_n)$$

which completes the proof. □

Lemma 4.2 (discrete to nonlocal control)**.** *Let $p \ge 1$. Assume Ω, μ, η, and x_i satisfy **(A1) - (A8)**. Let graph weights W_{ij} be given by* (1)*. Let constants $a, b > 0$ be such that for $\tilde\eta(|x|) = a$ for $|x| \le b$ and $\tilde\eta(|x|) = 0$ otherwise it holds that $\tilde\eta \le \eta$. Let T_n be a transport map satisfying the results of Theorem 3.3 and let $\tilde\varepsilon_n = \varepsilon_n - \frac{2\|T_n-\mathrm{Id}\|_{L^\infty}}{b}$. Then there exists constants $n_0 > 0$ and $C > 0$ (independent of n and f_n) such that for all $n \ge n_0$*

$$\mathcal{E}_{\tilde\varepsilon_n}^{(NL,p)}(f_n \circ T_n; \tilde\eta) \le C\mathcal{E}_n^{(p)}(f_n; \eta)$$

where $\mathcal{E}_{\tilde\varepsilon_n}^{(NL,p)}$ is defined by

$$\mathcal{E}_\varepsilon^{(NL,p)}(f;\eta) = \frac{1}{\varepsilon^p}\int_\Omega\int_\Omega \eta_\varepsilon(|x-z|)|f(x)-f(z)|^p\,\mathrm{d}x\,\mathrm{d}z. \tag{10}$$

Proof. Assume $\left|\frac{x-z}{\tilde\varepsilon_n}\right| < b$ then

$$|T_n(x)-T_n(z)| \le 2\|T_n-\mathrm{Id}\|_{L^\infty} + |x-z| \le 2\|T_n-\mathrm{Id}\|_{L^\infty} + b\tilde\varepsilon_n = b\varepsilon_n.$$

So,

$$\left|\frac{x-z}{\tilde\varepsilon_n}\right| < b \Rightarrow \left|\frac{T_n(x)-T_n(z)}{\varepsilon_n}\right| \le b$$

and therefore

$$\tilde\eta\left(\frac{|x-z|}{\tilde\varepsilon_n}\right) \le \tilde\eta\left(\frac{|T_n(x)-T_n(z)|}{\varepsilon_n}\right) \le \eta\left(\frac{|T_n(x)-T_n(z)|}{\varepsilon_n}\right).$$

Now,

$$\begin{aligned}\mathcal{E}_{\tilde{\varepsilon}_n}^{(NL,p)}(f_n \circ T_n) &\leq \frac{\varepsilon_n^d}{\tilde{\varepsilon}_n^{d+p}} \int_{\Omega^2} \eta_{\varepsilon_n}(|T_n(x) - T_n(z)|)\, |f_n(T_n(x)) - f_n(T_n(z))|^p \, \mathrm{d}x \, \mathrm{d}z \\ &= \frac{\varepsilon_n^{d+p}}{\left(\inf_{x\in\Omega} \rho^2(x)\right)\tilde{\varepsilon}_n^{d+p}} \mathcal{E}_n^{(p)}(f_n).\end{aligned}$$

Since $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$ we are done. □

In the next lemma we show that that boundedness of non-local energies implies regularity at scales greater ε. This allows us to relate non-local bounds to local bounds after mollification.

Lemma 4.3 (nonlocal to averaged local)**.** *Assume $\Omega \subset \mathbb{R}^d$ is open and bounded and $p \geq 1$. Assume that $\eta : [0,\infty) \to [0,\infty)$ is non-increasing, $\eta(0) > 0$ and η is continuous near 0. Then there exists a constant $C \geq 1$ and a mollifier J with $\operatorname{supp}(J) \subseteq \overline{B(0,1)}$ such that for all $\varepsilon > 0$, $f \in L^p(\Omega)$, $\Omega' \subset\subset \Omega$ with $\operatorname{dist}(\Omega', \partial\Omega) > \varepsilon$ it holds that*

$$\mathcal{E}_\infty^{(p)}(J_\varepsilon * f; \Omega') \leq C\mathcal{E}_\varepsilon^{(NL,p)}(f).$$

where $\mathcal{E}_\infty^{(p)}$ is defined by (5) *and $\mathcal{E}_\varepsilon^{(NL,p)}$ is defined by* (10)*.*

Proof. Let J be a radially symmetric mollifier supported in $B(0,1)$ and such that for some $\beta > 0$, $J \leq \beta\eta$ and $|\nabla J| \leq \beta\eta$. Without loss of generality we can assume $\operatorname{supp}(\eta) \subset \overline{B(0,1)}$. Let $g_\varepsilon = J_\varepsilon * f$. For arbitrary $x \in \Omega$ with $\operatorname{dist}(x, \partial\Omega) > \varepsilon$ we have

$$\begin{aligned}|\nabla g_\varepsilon(x)| &= \left| \int_\Omega \nabla J_\varepsilon\,(x-z)\, f(z)\, \mathrm{d}z \right| \\ &= \left| \int_\Omega \nabla J_\varepsilon\,(x-z)\,(f(z) - f(x))\, \mathrm{d}z - \int_{\mathbb{R}^d\setminus\Omega} \nabla J_\varepsilon\,(x-z)\, f(x)\, \mathrm{d}z \right| \\ &\leq \frac{\beta}{\varepsilon^{d+1}} \int_\Omega \eta\left(\frac{x-z}{\varepsilon}\right) |f(z) - f(x)|\, \mathrm{d}z + \frac{1}{\varepsilon^{d+1}} \int_{\mathbb{R}^d\setminus\Omega} \left|(\nabla J)\left(\frac{x-z}{\varepsilon}\right)\right| |f(x)|\, \mathrm{d}z.\end{aligned}$$

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

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

since for all $z \in \mathbb{R}^d \setminus \Omega$ 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,

$$\begin{aligned}|\nabla g_\varepsilon(x)|^p &\leq \beta^p \left(\int_\Omega \frac{1}{\varepsilon} \eta_\varepsilon(x-z)\, |f(z) - f(x)|\, \mathrm{d}z \right)^p \\ &\leq \gamma_\eta^{p-1} \beta^p \int_\Omega \eta_\varepsilon(x-z) \frac{|f(z) - f(x)|^p}{\varepsilon^p}\, \mathrm{d}z\end{aligned}$$

by Jensen's inequality and where $\gamma_\eta = \int_{B(0,1)} \eta(w)\, \mathrm{d}w$. Hence,

$$\begin{aligned}\int_{\Omega'} |\nabla g_\varepsilon(x)|^p \, \mathrm{d}x &\leq \gamma_\eta^{p-1} \beta^p \int_\Omega \int_\Omega \eta_\varepsilon(|x-z|) \left| \frac{f(z) - f(x)}{\varepsilon^p} \right|^p \mathrm{d}z \, \mathrm{d}x \\ &\leq \gamma_\eta^{p-1} \beta^p \mathcal{E}_\varepsilon^{(NL,p)}(f)\end{aligned}$$

which completes the proof. □

We prove the compactness property for bounded sequences. The convergence of a subsequence is a consequence of being able to bound $\tilde{g}_n = J_{\varepsilon_n} * (f_n \circ T_n)$ in $W^{1,p}$ (hence the sequence $\{\tilde{g}_n\}_n$ is precompact in $L^p(\mu)$) and show $\|f_n \circ T_n - \tilde{g}_n\|_{L^p} \to 0$.

Proposition 4.4 (compactness)**.** *Consider the assumptions and the graph construction of Lemma 4.1. Then with probability one, any sequence $f_n : \Omega_n \to \mathbb{R}$ with $\sup_{n\in\mathbb{N}} \mathcal{E}_n^{(p)}(f_n) < \infty$ and $\sup_{n\in\mathbb{N}} \|f_n\|_{L^\infty(\mu_n)} < \infty$ has a subsequence f_{n_m} such that (μ_{n_m}, f_{n_m}), converges in TL^p to (μ, f) for some $f \in L^p(\mu)$.*

Proof. Since $\mathcal{E}_n^{(p)}(f_n) \geq C\mathcal{E}_n^{(1)}(f_n)$ the compactness in TL^1 follows from Theorem 1.2 in [30]. We note that from the proof of Theorem 1.2 it follows that there in fact exists a subsequence f_{n_m}, and a sequence of transportation maps $T_{n_m \sharp}\mu = \mu_{n_m}$ such that

$$\lim_{m\to\infty} \|f - f_{n_m} \circ T_{n_m}\|_{L^1(\mu)} + \|T_{n_m} - \mathrm{Id}\|_{L^\infty(\mu)} = 0.$$

Since $\|f - f_{n_m} \circ T_{n_m}\|_{L^\infty(\mu)} < \infty$ the convergence of f_{n_m} to f in TL^p follows by interpolation. □

Lemma 4.5 (uniform convergence)**.** *Consider the assumptions and the graph construction of Lemma 4.1. Assume that $\varepsilon_n^p n \to 0$ as $n \to \infty$, which, due to **(A5)**, implies that $p > d$. Furthermore assume that with probability one $(\mu_n, f_n) \to (\mu, f)$ in TL^p metric as $n \to \infty$ and that $\sup_{n\in\mathbb{N}} \mathcal{E}_n^{(p)}(f_n) < \infty$. Then $f \in C^{0,\gamma}(\Omega)$, with $\gamma = 1 - \frac{d}{p} > 0$, and for all $\Omega' \subset\subset \Omega$*

$$\max_{\{k\,:\,x_k\in\Omega'\}} |f(x_k) - f_n(x_k)| \to 0 \qquad \text{as } n \to \infty.$$

Moreover, if for all $k = 1, \ldots, N$, $f_n(x_k) = y_k$ for all n, it follows that $f(x_k) = y_k$.

Proof. Find constants $a, b > 0$ such that $\tilde{\eta}(t) := a$ if $|t| \leq b$ and $\tilde{\eta}(t) := 0$ if $|t| > b$ satisfies $\tilde{\eta} \leq \eta$. Now we define $\tilde{f}_n = f_n \circ T_n$ where T_n is the transportation map satisfying the conclusions on Theorem 3.3 and set $\tilde{\varepsilon}_n = \varepsilon_n - \frac{2\|T_n - \mathrm{Id}\|_{L^\infty}}{b}$. Then for n sufficiently large $\tilde{\varepsilon}_n > 0$, and $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$. We note that if $|T_n(x) - T_n(z)| > b\varepsilon_n$ then

$$|x - z| \geq |T_n(x) - T_n(z)| - 2\|T_n - \mathrm{Id}\|_{L^\infty} > b\varepsilon_n - 2\|T_n - \mathrm{Id}\|_{L^\infty} = \tilde{\varepsilon}_n b.$$

Hence, $\tilde{\eta}\left(\frac{|x-z|}{\tilde{\varepsilon}_n}\right) \leq \tilde{\eta}\left(\frac{|T_n(x)-T_n(z)|}{\varepsilon_n}\right)$. Let $\mathcal{E}_{\tilde{\varepsilon}}^{(NL,p)}$ be the non-local Dirichlet energy defined in (10) with $\varepsilon = \tilde{\varepsilon}_n$ and $\eta = \tilde{\eta}$. Then, by Lemma 4.2

$$\mathcal{E}_{\tilde{\varepsilon}}^{(NL,p)}(\tilde{f}_n) \leq C\mathcal{E}_n^{(p)}(f_n).$$

Hence, $\mathcal{E}_{\tilde{\varepsilon}}^{(NL,p)}(\tilde{f}_n)$ is bounded and therefore, by Lemma 4.3 we have that $\mathcal{E}_\infty^{(p)}(J_{\tilde{\varepsilon}_n} * f_n; \Omega')$ is bounded for every $\Omega' \subset\subset \Omega$. One can easily show $\|J_{\tilde{\varepsilon}_n} * f_n\|_{L^p(\Omega')} \leq \|\tilde{f}_n\|_{L^p}$ and therefore $J_{\tilde{\varepsilon}_n} * \tilde{f}_n$ is locally bounded in $W^{1,p}$. We also note that since $f_n \circ T_n$ converges to f in $L^p(\mu)$

$$\begin{aligned}\|J_{\tilde{\varepsilon}_n} * \tilde{f}_n - f\|_{L^p(\Omega')} &\leq \|J_{\tilde{\varepsilon}_n} * \tilde{f}_n - J_{\tilde{\varepsilon}_n} * f + J_{\tilde{\varepsilon}_n} * f - f\|_{L^p(\Omega')} \\ &\leq \|\tilde{f}_n - f\|_{L^p(\Omega)} + \|J_{\tilde{\varepsilon}_n} * f - f\|_{L^p(\Omega')} \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

Since $J_{\tilde{\varepsilon}_n} * \tilde{f}_n \to f$ in $L^p(\Omega')$, by the compactness of the embedding of $W^{1,p}(\Omega')$ into $C^{0,\gamma}$ (Morrey's inequality), for $\gamma = 1 - \frac{d}{p}$, we have that

$$J_{\tilde{\varepsilon}_n} * \tilde{f}_n \to f \quad \text{ uniformly on } \Omega' \text{ as } n \to \infty.$$

Therefore, for each $k \in \{1, \dots, N\}$, $J_{\tilde{\varepsilon}_n} * \tilde{f}_n$ converges uniformly to f on $B(0, \delta)$ for any δ such that $B(x_k, \delta) \subset \Omega$. For any $x \in B(x_k, 3\tilde{\varepsilon}_n) \cap \Omega_n$ we have (for a constant C)

$$|f_n(x_k) - f_n(x)| \leq \operatorname{osc}_{3\tilde{\varepsilon}_n}(f_n)(x_k) \leq \operatorname{osc}_{4\varepsilon_n}(f_n)(x_k) \leq \left(4^p C \mathcal{E}_n^{(p)}(f_n) n \varepsilon_n^p\right)^{\frac{1}{p}} \to 0$$

by Lemma 4.1. It follows that

$$\max_{k=1,\dots,n} \max_{x \in B(x_k, 3\tilde{\varepsilon}_n) \cap \Omega_n} |f_n(x) - f_n(x_k)| \to 0.$$

To complete the proof we notice that for any $\Omega' \subset\subset \Omega$

$$\begin{aligned}
&\max_{\{k \,:\, x_k \in \Omega'\}} |f(x_k) - f_n(x_k)| \\
&\leq \max_{\{k \,:\, x_k \in \Omega'\}} |f(x_k) - J_{\tilde{\varepsilon}_n} * \tilde{f}_n(x_k)| + |J_{\tilde{\varepsilon}_n} * \tilde{f}_n(x_k) - f_n(x_k)| \\
&\leq \|f - J_{\tilde{\varepsilon}_n} * \tilde{f}_n\|_{L^\infty(\Omega')} + \max_{\{k \,:\, x_k \in \Omega'\}} \int_{B(0, 2\tilde{\varepsilon}_n)} J_{\tilde{\varepsilon}_n}(x_k - x) \, |f_n(T_n(x)) - f_n(x_k)| \, \mathrm{d}x \\
&\leq \|f - J_{\tilde{\varepsilon}_n} * \tilde{f}_n\|_{L^\infty(\Omega')} + \max_{\{k \,:\, x_k \in \Omega'\}} \sup_{x \in B(x_k, 3\tilde{\varepsilon}_n) \cap \Omega_n} |f_n(x) - f_n(x_k)|
\end{aligned}$$

and the above converges to zero for all x_k. □

4.1 Asymptotic consistency via Γ-convergence

We approach proving Theorem 2.1 using Γ-convergence. Namely as pointed out in Section 3.1 convergence of minimizers follows from Γ-convergence and compactness. We use the general setup of [30]. In particular we first establish in Lemma 4.6 that nonlocal functionals $\mathcal{E}_{\varepsilon_n}^{(NL,p)}$ Γ-converge to $\mathcal{E}_\infty^{(p)}$. We then state and prove the Γ-convergence of $\mathcal{E}_{n,con}^{(p)}$ towards $\mathcal{E}_\infty^{(p)}$ or $\mathcal{E}_{\infty,con}^{(p)}$ depending on how quickly $\varepsilon_n \to 0$ as $n \to \infty$. Steps of proving this claim rely on Lemma 4.6.

Lemma 4.6 (continuum nonlocal to local)**.** *Let $p > 1$. Assume Ω satisfy the assumptions **(A1) - (A2)** and η satisfies assumptions **(A6) - (A8)**. Then $\mathcal{E}_\varepsilon^{(NL,p)}$, defined in* (10)*, Γ-converges as $n \to \infty$ in $L^p(\Omega)$ to the functional $\mathcal{E}_\infty^{(p)}$ defined in* (5)*.*

If ρ is constant and Ω is convex this result is contained in the appendix to [3]. For general Ω it follows from Theorem 8 in [42]. We remark that while the functional in [42] appears different the term $|x - y|^p$ which arises can be absorbed in the kernel. The results can be extended to general ρ in a straightforward manner as has been done for $p = 1$ in Section 4 of [30] and has been remarked in Proposition 1.10 in [31].

Theorem 4.7 (discrete to local Γ-convergence)**.** *Let $p > 1$. Assume Ω, μ, η, and x_i satisfy the assumptions **(A1) - (A8)**. Let graph weights W_{ij} be given by* (1)*. Let $M \geq \max_{i=1,\dots,N} |y_i|$. Then with probability one $\mathcal{E}_{n,con}$, defined in* (3)*, Γ-converges as $n \to \infty$ in TL^p metric on the set $\{(\nu, g) : \nu \in \mathcal{P}(\Omega), \|g\|_{L^\infty(\nu)} \leq M\}$ to the functional*

$$\begin{cases} \mathcal{E}_{\infty,con}^{(p)} & \text{if } \lim_{n\to\infty} n\varepsilon_n^p = 0 \\ \mathcal{E}_\infty^{(p)} & \text{if } \lim_{n\to\infty} n\varepsilon_n^p = \infty \end{cases}$$

where $\mathcal{E}_\infty^{(p)}$ is defined in (5) *and $\mathcal{E}_{\infty,con}^{(p)}$ is defined in* (6)*.*

Restricting the space to the set of functions bounded by M is needed since the functional $\mathcal{E}_\infty^{(p)}$ is invariant under adding a constant and that due to the loss of constraints in the limit when $\lim_{n\to\infty} n\varepsilon_n^p = \infty$ the compactness needed does not hold. We note that placing an upper bound on f is not restrictive in practice since both discrete and continuum minimizers satisfy the bound.

We prove the liminf inequalities and the existence of a recovery sequence separately. Since $\mathcal{E}_\infty^{(p)} \le \mathcal{E}_{\infty,con}^{(p)}$ the liminf inequalities needed can be stated in the following way.

Lemma 4.8. *Under the same conditions as Theorem 4.7, with probability one, for any $f \in L^p$ with $\|f\|_{L^\infty(\mu)} \le M$ and any sequence $f_n \to f$ in TL^p with $\|f_n\|_{L^\infty(\mu_n)} \le M$ we have*

$$\mathcal{E}_\infty^{(p)}(f) \le \liminf_{n\to\infty} \mathcal{E}_n^{(p)}(f_n) \le \liminf_{n\to\infty} \mathcal{E}_{n,con}^{(p)}(f_n). \tag{11}$$

Furthermore if $\lim_{n\to\infty} n\varepsilon_n^p = 0$ then

$$\mathcal{E}_{\infty,con}^{(p)}(f) \le \liminf_{n\to\infty} \mathcal{E}_{n,con}^{(p)}(f_n). \tag{12}$$

Proof. Let $f_n \to f$ in TL^p. The first inequality of (11) follows from Lemma 4.6 in the same way the analogous result is shown for $p = 1$ in Section 5 of [30]. The second inequality follows from definition of $\mathcal{E}_n^{(p)}$ and $\mathcal{E}_{n,con}^{(p)}$

When $\lim_{n\to\infty} n\varepsilon_n^p = 0$ the inequality (12) is a consequence of Lemma 4.5. □

We now prove the existence of a recovery sequence. Since $\mathcal{E}_\infty^{(p)} \le \mathcal{E}_{\infty,con}^{(p)}$ we state it in the following way.

Lemma 4.9. *Under the same conditions as Theorem 4.7, with probability one, for any function $f \in L^p$, with $\|f\|_{L^\infty(\mu)} \le M$ there exists a sequence f_n satisfying $f_n \to f$ in TL^p with $\|f_n\|_{L^\infty(\mu_n)} \le M$ and*

$$\mathcal{E}_{\infty,con}^{(p)}(f) \ge \limsup_{n\to\infty} \mathcal{E}_{n,con}^{(p)}(f_n). \tag{13}$$

Furthermore if $\lim_{n\to\infty} n\varepsilon_n^p = \infty$ then

$$\mathcal{E}_\infty^{(p)}(f) \ge \limsup_{n\to\infty} \mathcal{E}_{n,con}^{(p)}(f_n). \tag{14}$$

Proof. The proof of the first inequality is a straightforward adaptation of the analogous result for $p = 1$ in Section 5 of [30]. The recovery sequence used is defined as a restriction of f to Ω_n: $f_n(x_i) = f(x_i)$ for all $i = 1, \dots, n$, and thus satisfies the constraints and $\|f_n\|_{L^\infty(\mu_n)} \le M$.

The same argument and recovery sequence construction can be used to show that with probability one, for any function $f \in L^p$, with $\|f\|_{L^\infty(\mu)} \le M$ there exists a sequence f_n satisfying $f_n \to f$ in TL^p with $\|f_n\|_{L^\infty(\mu_n)} \le M$ and

$$\mathcal{E}_\infty^{(p)}(f) \ge \limsup_{n\to\infty} \mathcal{E}_n^{(p)}(f_n). \tag{15}$$

Let us now consider that case that $n^p\varepsilon \to \infty$ as $n \to \infty$ and show the second inequality. Suppose $\mathcal{E}_{\infty,con}^{(p)}(f) < \infty$ else the lemma is trivial. Let f_n be the recovery sequence for (15).

We define $\hat{f}_n : \Omega_n \to \mathbb{R}$ by

$$\hat{f}_n(x_i) = \begin{cases} y_i & \text{for } i = 1, \dots, N, \\ f_n(x_i) & \text{for } i = N+1, \dots, n. \end{cases}$$

We note that $\hat{f}_n \to f$ in TL^p with $\|\hat{f}_n\|_{L^\infty(\mu_n)} \leq M$. To show (14) it suffices to show that

$$\lim_{n\to\infty} \mathcal{E}_n^{(p)}(f_n) - \mathcal{E}_{n,con}^{(p)}(\hat{f}_n) = 0. \tag{16}$$

We may write,

$$\begin{aligned}\left|\mathcal{E}_n^{(p)}(f_n) - \mathcal{E}_{n,con}^{(p)}(\hat{f}_n)\right| &\leq \frac{1}{\varepsilon_n^p}\frac{2}{n^2}\sum_{i=1}^{N}\sum_{j=1}^{n}\eta_{\varepsilon_n}(|x_i - x_j|)\,\big||f(x_i) - f(x_j)|^p - |y_i - f(x_j)|^p\big| \\ &\leq \frac{2^{p+1}M^p}{\varepsilon_n^p n}\sum_{i=1}^{N}\frac{1}{n}\sum_{j=1}^{n}\eta_{\varepsilon_n}(|x_i - x_j|)\end{aligned} \tag{17}$$

Step 1. Let us consider first the case that $\eta(t) = a$ if $|t| < b$ and $\eta(t) = 0$ otherwise for some $a, b > 0$. Then, using Theorem 3.3

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\eta_{\varepsilon_n}(|x_i - x_j|) &\leq \frac{\eta(0)}{\varepsilon^d}\mu_n(B(x_i, \varepsilon b) \\ &\leq \frac{\eta(0)}{\varepsilon^d}\mu(B(x_i, \varepsilon b + \|\mathrm{Id} - T_n\|_{L^\infty})) \\ &\leq \eta(0)\left(\frac{\varepsilon b + \|\mathrm{Id} - T_n\|_{L^\infty}}{\varepsilon}\right)^d \mathrm{Vol}(B(0,1))\|\rho\|_{L^\infty} \leq C.\end{aligned}$$

Combining this inequality with (17) implies (16).
Step 2. Consider now general η satisfying **(A6)-(A8)**. Let

$$\tilde{\eta}(t) = \begin{cases}\eta(0) & \text{if } |t| \leq 1 \\ \eta(t) & \text{otherwise.}\end{cases}$$

Note that $\tilde{\eta}$ is radially nonincreasing, $\tilde{\eta} \geq \eta$, and that $\tilde{\eta}((|x| - 1)_+) \leq \tilde{\eta}(|x|/2)$. Theorem 3.3 implies that for n large $\|\mathrm{Id} - T_n\|_{L^\infty} \leq \varepsilon_n$. Consequently

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\eta_{\varepsilon_n}(|x_i - x_j|) &\leq \frac{1}{n}\sum_{j=1}^{n}\tilde{\eta}_{\varepsilon_n}(|x_i - x_j|) \\ &= \frac{1}{\varepsilon_n^d}\int_\Omega \tilde{\eta}\left(\frac{|x_i - T_n(y)|}{\varepsilon_n}\right)d\mu(y) \\ &\leq \frac{1}{\varepsilon_n^d}\int_\Omega \tilde{\eta}\left(\frac{|x_i - y|}{2\varepsilon_n}\right)d\mu(y) \leq C\end{aligned}$$

where the penultimate inequality follows from $\frac{|x_i - T_n(y)|}{\varepsilon_n} \geq \left(\frac{|x_i - y| - \|T_n - \mathrm{Id}\|_{L^\infty}}{\varepsilon_n}\right)_+ \geq \left(\frac{|x_i - y|}{\varepsilon_n} - 1\right)_+$. Again combining this estimate with (17) implies (16). □

We now state the Γ-convergence result relevant for the penalized model $\mathcal{S}_n^{(p)}$.

Lemma 4.10. *Under the conditions of Proposition 2.2 we have:*

- *(compactness) Any sequence* $f_n : \Omega_n \to \mathbb{R}$ *with* $\sup_{n\in\mathbb{N}} \mathcal{S}_n^{(p)}(f_n) + \|f_n\|_{L^\infty(\mu_n)} < \infty$ *has, with probability one, a subsequence* f_{n_m} *such that there exists* $f_\infty \in W^{1,p}$ *with* $f_{n_m} \to f_\infty$ *in* TL^p.
- *(*Γ*-convergence, well-posed regime) If* $\varepsilon_n^p n \to 0$ *then, with probability one, on the set* (μ_n, f_n) *with* $\|f_n\|_{L^\infty(\mu_n)} \le M$*,*

$$\Gamma\text{-}\lim_{n\to\infty} \left(\mathcal{E}_n^{(p)} + \lambda R^{(q)}\right) = \mathcal{E}_\infty^{(p)} + \lambda R^{(q)}$$

where the Γ*-convergence is considered in* TL^p *topology.*

- *(*Γ*-convergence, degenerate regime) If* $\varepsilon_n^p n \to \infty$ *then, with probability one, on the set* (μ_n, f_n) *with* $\|f_n\|_{L^\infty(\mu_n)} \le M$*,*

$$\Gamma\text{-}\lim_{n\to\infty} \left(\mathcal{E}_n^{(p)} + \lambda R^{(q)}\right) = \mathcal{E}_\infty^{(p)},$$

where the Γ*-convergence is considered in* TL^p *topology.*

Proof. The compactness follows directly from Proposition 4.4.

When $\varepsilon_n^p n \to 0$, for the liminf inequality assume $f_n \to f$ in TL^p and $\liminf_{n\to\infty} \mathcal{E}_n^{(p)}(f_n) < \infty$. Then by Lemma 4.5 $f_n(x_k) \to f(x_k)$ for all $k \in \{1, \dots, N\}$ and hence $\lambda R^{(q)}(f_n) \to \lambda R^{(q)}(f)$. By (11) of Lemma 4.8 we have $\liminf_{n\to\infty} \left(\mathcal{E}_n^{(p)}(f_n) + \lambda R^{(q)}(f_n)\right) \ge \mathcal{E}_\infty^{(p)}(f) + \lambda R^{(q)}(f)$. The limsup inequality follows in a similar manner from equation 15 and Lemma 4.5.

If $\varepsilon_n^p n \to \infty$, then the liminf inequality follows from (11) of Lemma 4.8, while, the limsup inequality follows directly from

$$\limsup_{n\to\infty} \mathcal{E}_n^{(p)}(f_n) + \lambda R^{(q)}(f_n) \le \limsup_{n\to\infty} \mathcal{E}_{n,conn}^{(p)}(f_n) \le \mathcal{E}_\infty^{(p)}(f)$$

and Lemma 4.9. □

4.2 Proofs of Theorem 2.1 and Proposition 2.2

The Γ-convergence and compactness results above allow us to prove Theorem 2.1. It is a general result that Γ-convergence and compactness imply the convergence of minimizers (as well as of almost minimizers) to a minimizer of the limiting problem, see [7, Theorem 1.21] or Theorem 3.2.

Proof of Theorem 2.1. Let f_n be a minimizer of $\mathcal{E}_{n,con}^{(p)}$. Recall that $M \ge \|y\|_{L^\infty(\mu_n)}$. Note that if $\|f_n\|_{L^\infty(\mu_n)} > M$ then, since the graph is connected with high probability p_n, such that $\sum_{n=1}^\infty (1 - p_n) < \infty$, for $\hat{f}_n = (f_n \wedge M) \vee (-M)$ we have $\mathcal{E}_{n,con}^{(p)}(\hat{f}_n) < \mathcal{E}_{n,con}^{(p)}(f_n)$ which contradicts the definition of f_n. Thus with high probability $\|f_n\|_{L^\infty} \le M$ for each n, hence we can restrict the minimization to the set of (f_n, μ_n) such that $\|f_n\|_{L^\infty(\mu_n)} \le M$. This allows us to consider the setting of Theorem 4.7.

By compactness result of Proposition 4.4 there exists a subsequence f_{n_m} converging in TL^p to $f \in L^p(\mu)$.

To prove (i) assume that $n\varepsilon_n^p \to 0$ as $n \to \infty$. The uniform convergence of statement (a) then follows from Lemma 4.5. The Γ-convergence result of Theorem 4.7 implies that f minimizes $\mathcal{E}_{\infty,con}^{(p)}$. Since the minimizer of $\mathcal{E}_{\infty,con}^{(p)}$ is unique the convergence holds along the whole sequence, thus establishing statement (c).

To prove (ii) assume that $n\varepsilon_n^p \to 0$ as $n \to \infty$. Again, Theorem 4.7 implies that f minimizes $\mathcal{E}_\infty^{(p)}$. □

The results of the Proposition 2.2 are proved by the same arguments, with using Lemma 4.10 instead of Theorem 4.7.

5 Improved model

In Theorem 2.1 we proved that the model $\mathcal{E}^{(p)}_{n,con}$, defined in (3), is consistent as $n \to \infty$ only if

$$\frac{1}{n^p} \gg \varepsilon_n.$$

This upper bound is quite undesirable as it restricts the range of ε that can be used. Furthermore in a nonasymptotic regime, for large but fixed finite n, it provides no guidance to what ϵ are appropriate (small enough). Finally as our numerical experiments show, see Figures 1(a) and 2(a), the range of ϵ for which the limiting problem is approximated well can be quite narrow. This problem is particularly pronounced if $p > d$ is close to d, which is the regime identified in [14] as the most relevant for semi-supervised learning.

It would be advantageous to have another model, asymptotical consistent with $\mathcal{E}^{(p)}_{\infty,con}$ which would not require an upper bound on ε_n, other than $\varepsilon_n \to 0$ as $n \to \infty$. Here we introduce a new, related, model $\mathcal{F}^{(p)}_{n,con}$ which has the desired properties, and whose minimizers can be computed with the same algorithms as those for $\mathcal{E}^{(p)}_{n,con}$.

We define the set of functions which are constant near the labeled points:

$$C^{(\delta)}_n = \{f : \Omega_n \to \mathbb{R} \ : \ f(x_k) = y_i \ \text{ whenever } |x_k - x_i| < \delta \text{ for } i = 1, \dots, N\}$$

Let $L = \min\{|x_i - x_j| \ : \ i \neq j\}/2$ and $R_n = \min\{2\varepsilon_n, L\}$. The new functional is defined by

$$\mathcal{F}^{(p)}_{n,con}(f) = \begin{cases} \frac{1}{\varepsilon_n^p}\frac{1}{n^2}\sum_{i,j=1}^n W_{ij}|f(x_i) - f(x_j)|^p & \text{if } f \in C^{(R_n)}_n \\ \infty & \text{else.} \end{cases} \tag{18}$$

We note that for $f \in C^{(R_n)}_n$, $\mathcal{F}^{(p)}_{n,con}(f) = \mathcal{E}^{(p)}_{n,con}(f)$ and that $\mathcal{F}^{(p)}_{n,con}(f) \geq \mathcal{E}^{(p)}_{n,con}(f)$ for all f.

For the asymptotic consistency we still need to require $p > d$, since only then is the limiting model $\mathcal{E}^{(p)}_{\infty,con}$ well defined. In Theorem 2.1 this followed from the assumption $n\varepsilon_n^p \to 0$ as $n \to \infty$. Since we no longer require the upper bound on ε_n we need to require $p > d$ explicitly.

Theorem 5.1 (Consistency of the improved model)**.** *Let $p > d$. Assume Ω, μ, η, and x_i satisfy the assumptions **(A1) - (A8)**. Let graph weights W_{ij} be given by* (1)*. Let f_n be a sequence of minimizers of $\mathcal{F}^{(p)}_{n,con}$ defined in* (18)*. Then, almost surely, the sequence (μ_n, f_n) is precompact in the TL^p metric. The TL^p limit of any convergent subsequence, (μ_{n_m}, f_{n_m}), is of the form (μ, f) where $f \in W^{1,p}(\Omega)$ is a minimizer of $\mathcal{E}^{(p)}_{\infty,con}$ defined in* (6)*.*

Proof of the theorem is a straightforward modification of the proof of Theorem 2.1. It relies on the following Γ-convergence result.

Theorem 5.2 (discrete to local Γ-convergence)**.** *Let $M \geq \max_{i=1,\dots,N} |y_i|$. Under the conditions of Theorem 5.1, with probability one $\mathcal{F}_{n,con}$ Γ-converges as $n \to \infty$ in TL^p metric on the set $\{(\nu, g) \ : \ \nu \in \mathcal{P}(\Omega), \|g\|_{L^\infty(\nu)} \leq M\}$ to the functional $\mathcal{E}^{(p)}_{\infty,con}$.*

We note that statement (11) of Lemma 4.8, and Proposition 4.4 hold for $\mathcal{F}^{(p)}_{n,con}$ since $\mathcal{E}^{(p)}_{n,con} \leq \mathcal{F}^{(p)}_{n,con}$. We now turn to proving the liminf property and the existence of recovery sequence needed to show that $\mathcal{F}^{(p)}_{n,con}$ Γ converges in TL^p topology to $\mathcal{E}^{(p)}_{\infty,con}$.

Lemma 5.3. *Under the conditions of Theorem 5.1, with probability one, for any $f \in L^\infty(\mu)$ with $\|f\|_{L^\infty(\mu)} \leq M$ and any sequence $f_n \to f$ in TL^p with $\|f_n\|_{L^\infty(\mu_n)} \leq M$ we have*

$$\mathcal{E}^{(p)}_{\infty,con}(f) \leq \liminf_{n\to\infty} \mathcal{F}^{(p)}_{n,con}(f_n). \tag{19}$$

Proof. Consider a sequence f_n, uniformly bounded in $L^\infty(\mu_n)$ and convergent in TL^p and such that $\liminf_{n\to\infty} \mathcal{F}^{(p)}_{n,con}(f_n) < \infty$. Without a loss of generality we assume $\lim_{n\to\infty} \mathcal{F}^{(p)}_{n,con}(f_n) < \infty$. Note that in contrast to Lemma 4.8 we no longer require $n\varepsilon_n^p \to 0$ as $n \to \infty$. Therefore we can no longer use the uniform convergence of Lemma 4.5.

Nevertheless since for n large $f_n = y_i$ on $B(x_i, 2\varepsilon)$ and $\|\mathrm{Id} - T_n\|_{L^\infty} < \varepsilon$ we conclude that $\tilde{f}_n := f_n \circ T_n = y_i$ on $B(x_i, \varepsilon)$ and consequently that for $g_n := J_{\varepsilon_n} * \tilde{f}_n$ it holds that $g_n(x_i) = y_i$. Furthermore note that $\|g_n\|_{L^\infty} \leq M$. By bounds of Lemma 4.2 and Lemma 4.3, g_n is uniformly bounded in $W^{1,p}(\Omega')$ for any $\Omega' \subset\subset \Omega$. Arguing as in the proof of Lemma 4.5 we conclude that $g_n \to f$ in $L^p(\Omega)$. Since $p > d$, $W^{1,p}$ is compactly embedded in the space of continuous functions. This implies that g_n uniformly converges to f on sets compactly contained in Ω. Therefore $f(x_i) = y_i$ for all $i = 1, \ldots, N$. Combining this with statement (11) of Lemma 4.8 yields (19). □

Lemma 5.4. *Under the conditions of Theorem 5.1, with probability one, for any $f \in L^\infty(\mu)$ with $\|f\|_{L^\infty(\mu)} \leq M$ there exists a sequence $f_n \to f$ in TL^p with $\|f_n\|_{L^\infty(\mu_n)} \leq M$ such that*

$$\mathcal{E}^{(p)}_{\infty,con}(f) \geq \limsup_{n\to\infty} \mathcal{F}^{(p)}_{n,con}(f_n). \tag{20}$$

Proof. Assume $\|f\|_{L^\infty(\mu)} \leq M$ and $\mathcal{E}^{(p)}_{\infty,con}(f) < \infty$. Then $f \in W^{1,p}(\Omega)$ and since $p > d$, f is continuous. Furthermore $f(x_i) = y_i$ for all $i = 1, \ldots, N$.

If there exists $\delta > 0$ such that $f \in W^{1,p}(\Omega)$ satisfies $f(x) = y_i$ for all $x \in B(x_i, \delta)$ and $i = 1, \ldots, N$ then the proof of (20) is the same as the proof of (13). In particular one can use the restriction of f to data points to construct a recovery sequence.

To treat general f in $W^{1,p}(\Omega)$ it suffices to find a sequence $g_n \in W^{1,p}(\Omega)$ satisfying the conditions above, namely such that $\|g_n\|_{L^\infty} \leq M$, $g_n(x) = y_i$ for all $x \in B(x_i, \delta_n)$ for a sequence $\delta_n \geq R$ converging to zero, which satisfies

$$\lim_{n\to\infty} \mathcal{E}^{(p)}_{\infty,con}(g_n) = \mathcal{E}^{(p)}_{\infty,con}(f). \tag{21}$$

We construct the sequence in the following way. Let θ be a cut-off function supported in $B(0,2)$. That is assume $\theta : \mathbb{R}^d \to [0,1]$ is smooth, radially symmetric and nonincreasing such that $\theta = 1$ on $B(0,1)$, $\theta = 0$ outside of $B(0,2)$, and $|\nabla\theta| < 2$. Define $\theta_\delta(z) = \theta(z/\delta)$.

We first consider the case $N = 1$. Let

$$g_n(x) = (1 - \theta_{\delta_n}(x - x_1))f(x) + \theta_{\delta_n}(x - x_1)y_1.$$

Then

$$\left|\mathcal{E}^{(p)}_{\infty,con}(g_n) - \mathcal{E}^{(p)}_{\infty,con}(f)\right| \geq \sigma_n \int_\Omega \left||\nabla g_n|^p - |\nabla f|^p\right| \rho^2 \,\mathrm{d}x \leq \sigma_n \int_{B(0,2\delta_n)} \left(|\nabla g_n|^p + |\nabla f|^p\right) \rho^2 \,\mathrm{d}x$$

We estimate

$$\int_{B(0,2\delta_n)} |\nabla g_n|^p \rho^2 \,\mathrm{d}x \leq 2^p \int_{B(0,2\delta_n)} |(f(x_1) - f(x))\nabla\theta_{\delta_n}(x - x_1)|^p + |\nabla f(x)|^p \rho^2 \,\mathrm{d}x$$

Using that $f \in C^{0,1-d/p}$ and furthermore, by the remark following Theorem 4 in Section 5.6.2 of [18] we obtain

$$\begin{aligned}\int_{B(0,2\delta_n)} |(f(x) - f(x_1))\nabla\theta_{\delta_n}(x - x_1)|^p \rho^2(x) \,\mathrm{d}x &\leq C_1 \delta_n^{p-d} \|\nabla f\|^p_{L^p(B(0,2\delta_n))} \|\nabla\theta_{\delta_n}\|^p_{L^p(B(x_1,2\delta_n))} \\ &\leq C_1 \|\nabla f\|^p_{L^p(B(0,2\delta_n))} \|\nabla\theta\|^p_{L^p(\mathbb{R}^d)}.\end{aligned}$$

Since $\lim_{n\to\infty} \int_{B(0,4\delta_n)} |\nabla f(x)|^p dx = 0$, by combining the inequalities above we conclude that (21) holds.

Generalizing to $N > 1$ is straightforward. □

6 Numerical experiments

The results of Theorem 2.1 show that when $\varepsilon_n^p n \to \infty$ then the the SSL problem (3) converges, while when $\varepsilon_n^p n \to \infty$ it degenerates as $n \to \infty$. However, in practice, for finite n, this does not provide a precise guidance on what ε are appropriate. We investigate, via numerical experiments in 1D, the affect of ε on solutions to (3) in elementary examples. We also numerically compare the results with our improved model (18).

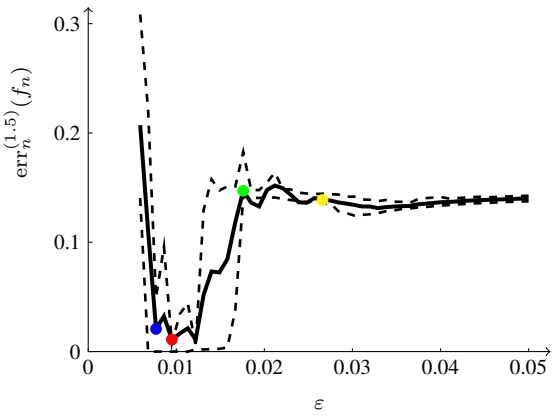
Let μ be the uniform measure on $[0, 1]$ and consider η defined by $\eta(t) = 1$ if $t \leq 1$ and $\eta(t) = 0$ otherwise. We choose two different values of p: $p = 1.5$ and $p = 2$. The training set is $\{(0, 0), (1, 1)\}$, that is we condition on functions f_n taking the value 0 at $x_1 = 0$ and taking the value 1 at $x_2 = 1$ (so $N = 2$). We avoid using $p = 1$ since any increasing function f with $f(0) = 0$ and $f(1) = 1$ is a minimizer to the limiting problem. For $p > 1$ the solution to the constrained limiting problem is $f^\dagger(x) = x$. Since $f^\dagger$ is continuous we can consider the following simple-to-compute notion of error: $\mathrm{err}_n^{(p)}(f_n) = \|f_n - f^\dagger\|_{L^p(\mu_n)}$.

To find minimizers of (3) we use coordinate gradient descent. The number of data points varies from $n = 50$ to $n = 10000$. For each n and each ε we consider 10 different realizations of the random sample and plot the average results. When ε is too small the graph is disconnected and we should not expect informative solutions, when ε is large we expect the averaging affect to cause degeneracy. On Figure 1(a) and Figure 2(a) we plot the error as a function of ϵ for fixed $n = 1000$. We see a clear regions where ε is too small and where ε is too large, with the intermediate range producing good estimators. Plots of minimizers for a particular ϵ in the "large-ε" region, show that they exhibit spikes, as expected.

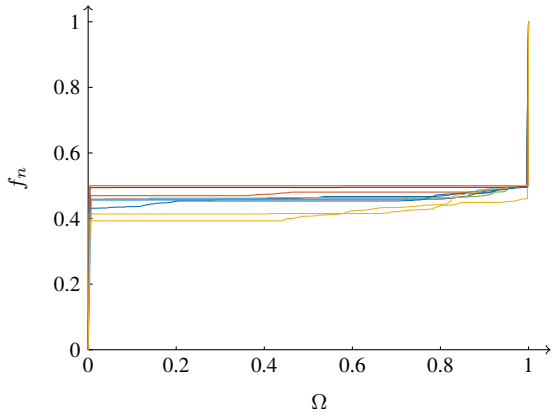
To measure how the transition point in ε where minimizers change behavior scale with n we define the following:

(i) $\varepsilon_{\mathrm{conn}}(n)$ is the smallest ε such that the graph with weights $W_{ij} = \eta_\varepsilon(|x_i - x_j|)$ is connected,

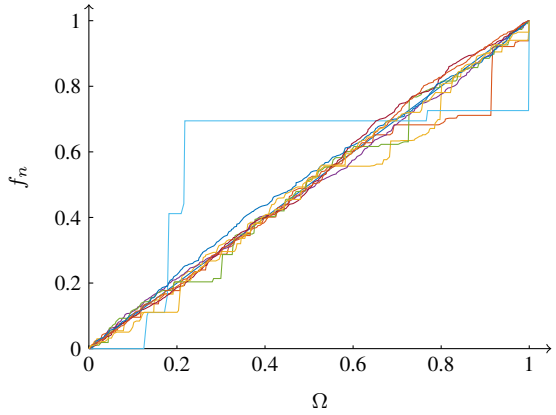
(ii) $\varepsilon_*^{(p)}(n)$ is the empirically best choice for ε, namely the ε that minimizes $\mathrm{err}_n^{(p)}(f_n)$ where f_n is the minimizer of (3) with $\varepsilon_n = \varepsilon$, and



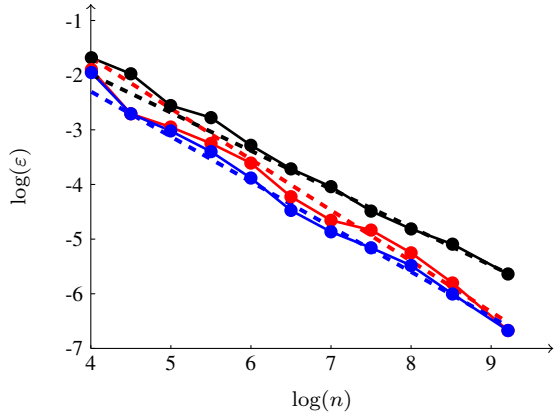
(a) Error of the function f_n output by the algorithm for $n = 1000$. The solid line is the mean error, the dashed lines are the 20% and 80% quantiles. We mark the connectivity bound $\varepsilon_{\mathrm{conn}}$ in blue, the optimal choice $\varepsilon_*^{(1.5)}$ in red and the upper bound $\varepsilon_{\mathrm{upper}}^{(1.5)}$ in green.



(b) We plot the functions output from the algorithm corresponding to multiple realisations of the data for $n = 1000$ and $\varepsilon = 0.0266$ (marked in yellow in Figure (a)).



(c) We plot the functions output from the algorithm corresponding to multiple realizations of the data for $n = 1000$ and $\varepsilon = \varepsilon_*^{(1.5)}$.

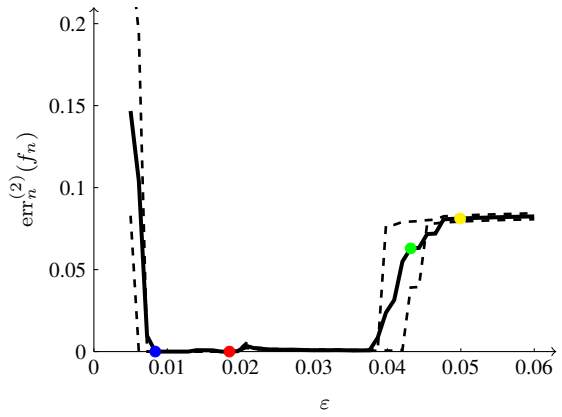


(d) Scaling in ε. The black line is $\varepsilon_{\mathrm{upper}}^{(1.5)}$, the red is $\varepsilon_*^{(1.5)}$, and the blue is $\varepsilon_{\mathrm{conn}}$. The dashed line indicates the best linear fit.

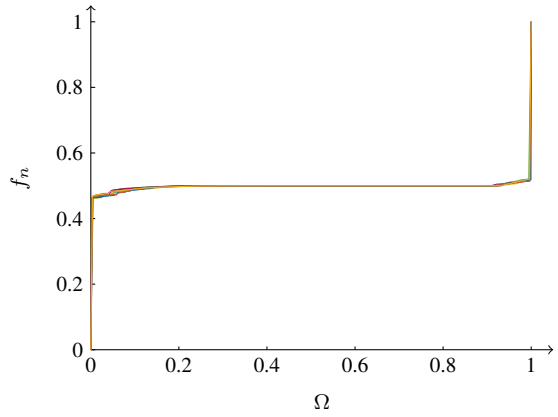
Figure 1: 1D Numerical Experiments averaged over 10 realizations for (3) with $p = 1.5$.

(iii) $\varepsilon_{\mathrm{upper}}^{(p)}(n)$ is the upper bound on ε for which the algorithm behaves well, which we identify as the maximizer of the second derivative of $-\mathrm{err}_n^{(p)}(f_n)$ with respect to ε, among $\varepsilon \geq \varepsilon_*^{(p)}(n)$. While computing $\varepsilon_{\mathrm{upper}}^{(p)}(n)$ we smooth the error slightly so that the method is robust to small perturbations.

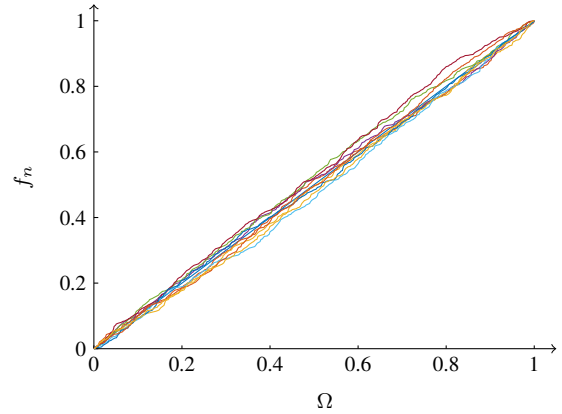
All of these points are highlighted on Figure 1(a) on Figure 2(a). In Figure 1(d) we plot how these values of ε scale with n. The best linear fit (based on five largest values of n) in the log-log domain



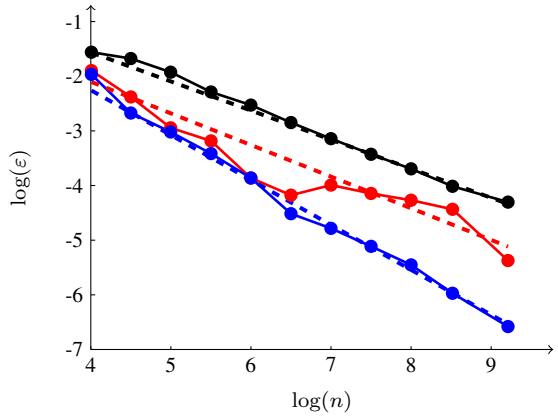
(a) Error of the function f_n output by the algorithm for $n = 1000$. The solid line is the mean error, the dashed lines are the 20% and 80% quantiles. We mark the connectivity bound $\varepsilon_{\text{conn}}$ in blue, the optimal choice $\varepsilon_*^{(2)}$ in red and the upper bound $\varepsilon_{\text{upper}}^{(2)}$ in green.



(b) We plot the functions output from the algorithm corresponding to multiple realisations of the data for $n = 1000$ and $\varepsilon = 0.05$ (marked in yellow in Figure (a)).



(c) We plot the functions output from the algorithm corresponding to multiple realisations of the data for $n = 1000$ and $\varepsilon = \varepsilon_*^{(2)}$.



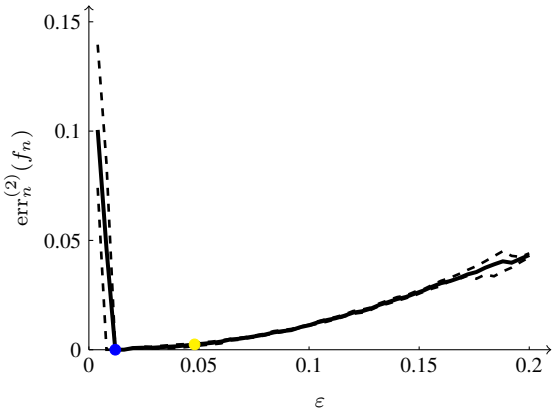
(d) Scaling in ε. The black line is $\varepsilon_{\text{upper}}^{(2)}$, the red is $\varepsilon_*^{(2)}$, and the blue is $\varepsilon_{\text{conn}}$. The dashed line indicates the best linear fit.

Figure 2: 1D Numerical Experiments averaged over 10 realizations for (3) with $p = 2$.

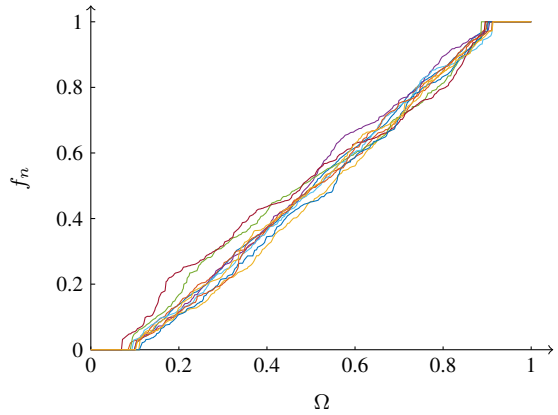
gives the following scalings

$$\varepsilon_*^{(1.5)} = \frac{7.678}{n^{0.930}} \qquad \varepsilon_{\text{upper}}^{(1.5)} = \frac{2.250}{n^{0.699}}$$
$$\varepsilon_*^{(2)} = \frac{1.240}{n^{0.579}} \qquad \varepsilon_{\text{upper}}^{(2)} = \frac{1.761}{n^{0.532}}$$
$$\varepsilon_{\text{conn}} = \frac{2.722}{n^{0.825}}.$$

We observe that our asymptotic scaling in $\varepsilon_{\text{upper}}^{(p)}$ is $\frac{1}{n^{0.5}}$ for $p = 2$ and $\frac{1}{n^{0.667}}$ for $p = 1.5$, which closely agrees with our numerical results. The true scaling in the connectivity of the graph is $\frac{\log(n)}{n}$,



(a) Error of the function f_n output by the algorithm for $n = 1000$. The solid line is the mean error, the dashed lines are the 20% and 80% quantiles. We mark the connectivity bound $\varepsilon_{\mathrm{conn}}$ in blue.



(b) We plot the functions output from the algorithm corresponding to multiple realisations of the data for $n = 1000$ and $\varepsilon = 0.05$ (marked in yellow in Figure (a)).

Figure 3: 1D Numerical Experiments averaged over 10 realizations for model (18) with $p = 2$.

our numerical results behave approximately as $\frac{1}{n^{0.825}}$. The difference is largely due to a small number of realizations (ten) we considered.

The optimal choice $\varepsilon_*^{(p)}$ does not fit as well to a linear function, this is most likely due to error, $\mathrm{err}_n^{(p)}$, being rather flat as a function of ε around the optimal value and hence there being large variability in $\varepsilon_*^{(p)}$.

The improved model (18), for which we show results in Figure 3, is far more robust to the choice of ε. We plot the error as a function of ε for $n = 1000$ and we see a much larger range in the admissible choices of ε. To highlight the difference we plot in Figure 3(b) outputs from multiple realizations of the data under the same conditions as for Figure 2(b), in particular we use the same data sequences and the same choice of ε. Note that the horizontal axis covers a much larger range on Figure 3(b). The comparison shows that model (3) does not produce a reasonable output, while all outputs of (18) are close to the truth.

Acknowledgements

The authors thank Matt Dunlop and Andrew Stuart for enlightening exchanges. This material is based on work supported by the National Science Foundation under the grants CCT 1421502 and DMS 1516677. The authors are also grateful to the Center for Nonlinear Analysis (CNA) for support.

References

[1] M. Ajtai, J. Komlós, and G. Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, 1984.

[2] M. Alamgir and U. Von Luxburg. Phase transition in the family of p-resistances. In *Advances in Neural Information Processing Systems 24*, pages 379–387, 2011.

[3] G. Alberti and G. Bellettini. A non-local anisotropic model for phase transitions: asymptotic behaviour of rescaled energies. *European J. Appl. Math.*, 9(3):261–284, 1998.

[4] M. Belkin and P. Niyogi. Using manifold structure for partially labeled classification. In *Advances in Neural Information Processing Systems (NIPS)*, 2003.

[5] M. Belkin and P. Niyogi. Semi-supervised learning on Riemannian manifolds. *Machine learning*, 56(1):209–239, 2004.

[6] M. Belkin and P. Niyogi. Convergence of Laplacian eigenmaps. In *Advances in Neural Information Processing Systems*, pages 129–136, 2007.

[7] A. Braides. Γ-*Convergence for Beginners*. Oxford University Press, 2002.

[8] T. Bühler and M. Hein. Spectral clustering based on the graph p-Laplacian. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 81–88, 2009.

[9] D. Burago, S. Ivanov, and Y. Kurylev. A graph discretization of the Laplace-Beltrami operator. *J. Spectr. Theory*, 4(4):675–714, 2014.

[10] R. R. Coifman and S. Lafon. Diffusion maps. *Appl. Comput. Harmon. Anal.*, 21(1):5–30, 2006.

[11] M. G. Crandall, L. C. Evans, and R. F. Gariepy. Optimal Lipschitz extensions and the infinity Laplacian. *Calculus of Variations and Partial Differential Equations*, 13(2):123–139, 2001.

[12] G. Dal Maso. *An Introduction to* Γ-*Convergence*. Springer, 1993.

[13] E. Davis and S. Sethuraman. Consistency of modularity clustering on random geometric graphs. *arXiv preprint arXiv:1604.03993*, 2016.

[14] A. El Alaoui, X. Cheng, A. Ramdas, M. J. Wainwright, and M. I. Jordan. Asymptotic behavior of ℓ_p-based Laplacian regularization in semi-supervised learning. In *29th Annual Conference on Learning Theory*, pages 879–906, 2016.

[15] A. Elmoataz, X. Desquesnes, and O. Lezoray. Non-local morphological PDEs and p-Laplacian equation on graphs with applications in image processing and machine learning. *IEEE Journal of Selected Topics in Signal Processing*, 6(7):764–779, 2012.

[16] A. Elmoataz, F. Lozes, and M. Toutain. Nonlocal pdes on graphs: From tug-of-war games to unified interpolation on images and point clouds. *Journal of Mathematical Imaging and Vision*, 57(3):381–401, 2017.

[17] A. Elmoataz, M. Toutain, and D. Tenbrinck. On the p-Laplacian and ∞-Laplacian on graphs with applications in image and data processing. *SIAM Journal on Imaging Sciences*, 8(4):2412–2451, 2015.

[18] L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 2010.

[19] C. Fefferman. Fitting a C^m-smooth function to data III. *Annals of Mathematics*, 2009.

[20] C. Fefferman, A. Israel, and G. K. Luli. Fitting a Sobolev function to data I. *Revista Matemática Iberoamericana*, 32(1):275–376, 2016.

[21] C. Fefferman, A. Israel, and G. K. Luli. Fitting a Sobolev function to data II. *Revista Matemática Iberoamericana*, 32(2):649–750, 2016.

[22] C. Fefferman, A. Israel, and G. K. Luli. Fitting a Sobolev function to data III. *Revista Matemática Iberoamericana*, 32(3):1039–1126, 2016.

[23] C. Fefferman and B. Klartag. Fitting a C^m-smooth function to data I. *Annals of mathematics*, 169(1):315–346, 2009.

[24] C. Fefferman and B. Klartag. Fitting a C^m-smooth function to data II. *Revista Matemática Iberoamericana*, 25(1):49–273, 2009.

[25] I. Fonseca and G. Leoni. *Modern Methods in the Calculus of Variations: L^p Spaces*. Springer Science & Business Media, 2007.

[26] N. García-Trillos. Variational limits of k-nn graph based functionals on data clouds. *arXiv preprint arXiv:1607.00696*, 2016.

[27] N. García Trillos, M. Gerlach, M. Hein, and D. Slepčev. Spectral convergence of the empirical graph Laplacian. *In Preparation*, 2017.

[28] N. García Trillos and R. Murray. A new analytical approach to consistency and overfitting in regularized empirical risk minimization. *arXiv preprint arXiv: 1607.00274*, 2016.

[29] N. García Trillos and D. Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *Canadian Journal of Mathematics*, 67:1358–1383, 2015.

[30] N. García Trillos and D. Slepčev. Continuum limit of Total Variation on point clouds. *Archive for Rational Mechanics and Analysis*, 220(1):193–241, 2016.

[31] N. García Trillos and D. Slepčev. A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis*, 2016.

[32] N. García Trillos, D. Slepčev, J. von Brecht, T. Laurent, and X. Bresson. Consistency of cheeger and ratio graph cuts. *Journal of Machine Learning Research*, 2015.

[33] E. Giné and V. Koltchinskii. Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 238–259. Inst. Math. Statist., Beachwood, OH, 2006.

[34] M. Hein. Uniform convergence of adaptive graph-based regularization. In *International Conference on Computational Learning Theory*, pages 50–64, 2006.

[35] M. Hein, J.-Y. Audibert, and U. von Luxburg. From graphs to manifolds–weak and strong pointwise consistency of graph Laplacians. In *Learning theory*, pages 470–485. Springer, 2005.

[36] T. Leighton and P. Shor. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica*, 9(2):161–187, 1989.

[37] Z. Li and Z. Shi. A convergent point integral method for isotropic elliptic equations on a point cloud. *Multiscale Modeling & Simulation*, 14(2):874–905, 2016.

[38] Z. Li, Z. Shi, and J. Sun. Point integral method for solving poisson-type equations on manifolds from point clouds with convergence guarantees. *Communications in Computational Physics*, 22(1):228–258, 2017.

[39] B. Nadler, N. Srebro, and X. Zhou. Statistical analysis of semi-supervised learning: The limit of infinite unlabelled data. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1330–1338, 2009.

[40] B. Pelletier and P. Pudlo. Operator norm convergence of spectral clustering on level sets. *J. Mach. Learn. Res.*, 12:385–416, 2011.

[41] M. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[42] A. C. Ponce. A new approach to Sobolev spaces and connections to Γ-convergence. *Calc. Var. Partial Differential Equations*, 19(3):229–255, 2004.

[43] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87. Springer, 2015.

[44] P. W. Shor and J. E. Yukich. Minimax grid matching and empirical measures. *Ann. Probab.*, 19(3):1338–1348, 1991.

[45] A. Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006.

[46] A. Singer and H.-T. Wu. Spectral convergence of the connection Laplacian from random samples. *Information and Inference: A Journal of the IMA*, 6(1):58–123, 2017.

[47] Michel Talagrand. *Upper and lower bounds of stochastic processes*, volume 60 of *Modern Surveys in Mathematics*. Springer-Verlag, Berlin Heidelberg, 2014.

[48] M. Thorpe, S. Park, S. Kolouri, G. K. Rohde, and D. Slepčev. A transportation L^p distance for signal analysis. *to appear in the Journal of Mathematical Imaging and Vision, arXiv preprint arXiv:1609.08669*, 2017.

[49] M. Thorpe and D. Slepčev. Transportation L^p distances: Properties and extensions. *In preparation*, 2017.

[50] M. Thorpe and F. Theil. Asymptotic analysis of the Ginzburg-Landau functional on point clouds. *to appear in the Proceedings of the Royal Society of Edinburgh Section A: Mathematics, arXiv preprint arXiv:1604.04930*, 2017.

[51] M. Thorpe, F. Theil, A. M. Johansen, and N. Cade. Convergence of the k-means minimization problem using Γ-convergence. *SIAM Journal on Applied Mathematics*, 75(6):2444–2474, 2015.

[52] D. Ting, L. Huang, and M. I. Jordan. An analysis of the convergence of graph Laplacians. In *Proceedings of the 27th International Conference on Machine Learning*, 2010.

[53] C. Villani. *Optimal Transport Old and New*. Springer-Verlag Berlin Heidelberg, 2009.

[54] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 2007.

[55] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, 2008.

[56] X. Wang. Spectral convergence rate of graph Laplacian. *arXiv preprint arXiv:1510.08110*, 2015.

[57] D. Zhou and B. Schölkopf. Regularization on discrete spaces. In *Proceedings of the 27th DAGM Conference on Pattern Recognition*, PR'05, pages 361–368, Berlin, Heidelberg, 2005. Springer-Verlag.

[58] X. Zhou and M. Belkin. 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*, pages 892–900, 2011.

[59] X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the 20th International Conference on Machine Learning*, pages 912–919, 2003.