ANALYSIS OF p-LAPLACIAN REGULARIZATION IN SEMI-SUPERVISED LEARNING *

DEJAN SLEPČEV † AND MATTHEW THORPE ‡

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 random geometric graphs, with connection radius $\varepsilon(n)$, to represent the geometry of the data set. Functionals which model the task reward the regularity of the estimator function and impose or reward the agreement with the training data. Here we consider discrete p-Laplacian regularization.

We investigate asymptotic behavior when the number of unlabeled points increases while the number of training points remains fixed. A delicate interplay between the regularizing nature of the functionals and the nonlocality inherent to the graph constructions is uncovered. Rigorous, almost optimal, ranges on the scaling of $\varepsilon(n)$ for asymptotic consistency are obtained and in these admissible ranges it is shown that minimizers of the discrete functionals converge uniformly to the desired continuum limit. Furthermore, we discover that, for the standard model, there is a restrictive upper bound on how quickly $\varepsilon(n)$ must converge to zero as $n \to \infty$. A new model is introduced, which is as simple as the original model, but overcomes this restriction.

Key words. semi-supervised learning, label propagation, regression, asymptotic consistency, Gamma-convergence, PDEs on graphs, nonlocal variational problems

AMS subject classifications. 49J55, 49J45, 62G20, 35J20, 65N12

1. Introduction.

Due to its applicability across a large spectrum of problems semi-supervised learning (SSL) is an important tool in data analysis. It deals with situations where one has access to relatively few labeled data points but potentially a large number of unlabeled data points. 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 $n - N$ unlabeled 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 by taking advantage of the information provided by the full data set $\Omega_n = \{x_i\}_{i=1}^n$. 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 represent the geometry we build a graph whose vertices are Ω_n 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 a decreasing function $\eta : [0, \infty) \to [0, \infty)$ with $\lim_{r \to \infty} \eta(r) = 0$. For a 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 label propagation problem is to find an estimator $u : \Omega_n \to \mathbb{R}$ which agrees with preassigned labels. To solve the regression problem one considers objective functionals which penalize the lack of smoothness of u and take the structure of the graph into account. In particular, we consider the functionals which generalize the graph Laplacian, namely the graph p-Laplacian. A particular objective functional we consider is

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

We consider minimizing $\mathcal{E}_n^{(p)}(f)$ under the constraint that

$$f(x_i) = y_i \ \text{ for all } \ i = 1, \dots, N. \tag{2}$$

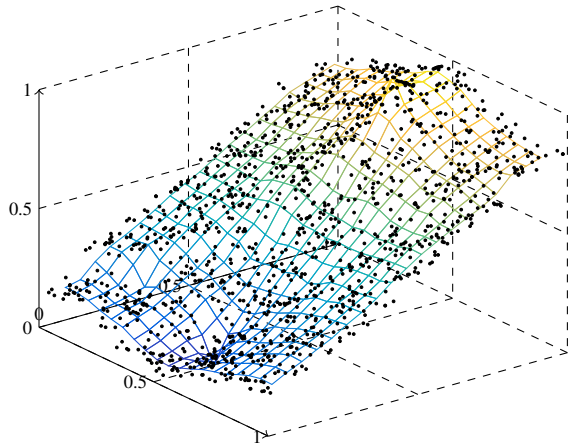
A numerically computed example of the minimizer of the functional is shown in Figure 1(a).

We investigate the asymptotic behavior in the limit when the number of unlabeled data points n goes to infinity while the number of training data points N is fixed. This is consistent with the semi-supervised learning paradigm of having few labeled points and an abundance of unlabeled data. As $n \to \infty$, $\varepsilon_n \to 0$ to increase the resolution

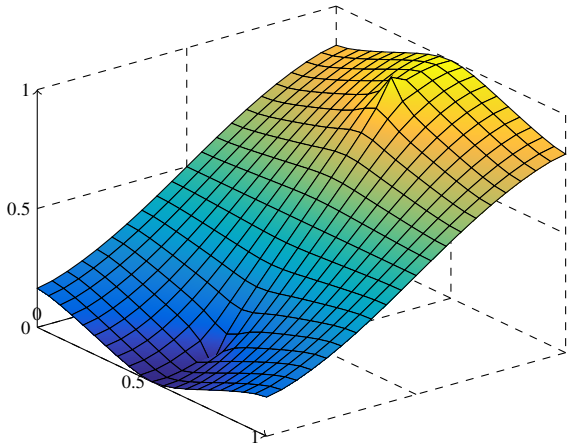
* This material is based on work supported by the National Science Foundation under the grants CCT 1421502, DMS 1516677, and DMS-1814991. The authors are also grateful to the Center for Nonlinear Analysis (CNA) for support. MT is grateful for the support of the Cantab Capital Institute for the Mathematics of Information at the University of Cambridge.

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, 15213, USA, slepcev@math.cmu.edu

‡Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, CB3 0WA, UK, m.thorpe@maths.cam.ac.uk



(a) A minimizer of (1) under constraint (2) for $\varepsilon = 0.058$ and $\eta = \mathbb{1}_{[0,1]}$ and $n = 1280$. The grid is to aid visualization.



(b) Minimizer of the continuum functional (3) under constraint (2).

Fig. 1: A 2D numerical experiment for measure μ with density one on $[0,1]^2$, training data $x_1 = (0.2, 0.5)$ and $x_2 = (0.8, 0.5)$ and labels $y_1 = 0$ and $y_2 = 1$, and $p = 4$.

and limit the computational cost. Namely, as ε_n is the length scale over which the information on μ is averaged, taking ε_n to zero ensures that the finer scales of μ are resolved as more data points become available. As regards the computational cost, for small ε_n and compactly supported η, the matrix (W_{ij}) of edge weights is sparse. More precisely the number of edges, is proportional to ε_n^d. The cost of computing the gradient of the functional grows linearly with the number of edges, as does the number of nonzero entries of the Hessian of the functional. This directly affects the complexity of numerical methods used.

While in this paper we consider data distributed in a set of full dimension, we remark that there are no essential difficulties to extend the results to the manifold setting, namely one where μ is a measure supported on compact manifold $\mathcal{M}$ of dimension d embedded in $\mathbb{R}^D$. Such extensions have already been done for related problems concerning the graph Laplacian [32], where the modification of background results (such as optimal transportation estimates) has been carried out.

The continuum limiting problem corresponds to minimizing

$$\mathcal{E}_\infty^{(p)}(f) = \sigma \int_\Omega |\nabla f(x)|^p \, \rho^2(x) \, \mathrm{d}x, \tag{3}$$

where σ is a constant that depends on η, subject to the constraint that $f(x_i) = y_i$ for $i = 1, \ldots, N$. A numerically computed minimizer of the functional is shown in Figure 1(b). 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 Sobolev embedding ensures that functions in $W^{1,p}$ are continuous. When $p \leq d$ and $d > 1$, one cannot expect to be able to impose pointwise data. Indeed, spikes were observed in discrete models with graph-Laplacian-based regularizations (that is for $p = 2$) by Nadler, Srebro and Zhou in [48] who also argued that they arise since there exist functions with arbitrarily small energy $\mathcal{E}_\infty^{(p)}(f)$, for $p = 2$, which agree with labels on the training set. In [19] El Alaoui, Cheng, Ramdas, Wainwright and Jordan go a step further and suggest $p = d$ as the transition point between the regime where spikes appear and where solutions are "smooth". They argue, based on the Sobolev embedding theorem, 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 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 verify the conclusions of [19]. More precisely, our aim was to show that constrained minimizers of $\mathcal{E}_n^{(p)}(f)$ converge as $n \to \infty$, in the appropriate topology, to minimizers of $\mathcal{E}_\infty^{(p)}(f)$, subject to constraints when $p > d$, and without constraints when $p \leq d$. However we discovered an additional phenomenon,

namely that the undesirable spikes in the minimizers to graph p-Laplacian can occur even when $p > d$. Namely, [19] shows pointwise convergence of the form

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \mathcal{E}_n^{(p)}(f) = \mathcal{E}_\infty^{(p)}(f),$$

when f is smooth enough. However considering a fixed function f is not sufficient to conclude that the constrained minimizers of $\mathcal{E}_n^{(p)}$ converge to constrained minimizers of $\mathcal{E}_\infty^{(p)}$. In fact answering that question requires a set of tools from applied analysis which we discuss below. We show, roughly speaking, that for $d \geq 3$ the convergence of minimizers holds if and only if

$$\left(\frac{1}{n}\right)^{\frac{1}{p}} \gg \varepsilon_n \gg \left(\frac{\ln n}{n}\right)^{\frac{1}{d}} \quad \text{as } n \to \infty. \tag{4}$$

The lower bound above is related to the connectivity of the graph and is well understood, [37, 38]. Our lower bounds for $d = 1, 2$ contain additional correction terms and are not optimal. Our upper bound implies that the models 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.

In the "ill-posed" regime $\varepsilon_n^p n \to \infty$, under the usual connectivity requirement (which when $d \geq 3$ reads $\varepsilon_n^d \frac{n}{\ln n} \to \infty$), we are still able to establish the asymptotic behavior of algorithms. Namely, we show that minimizers of $\mathcal{E}_n^{(p)}$ with constraints converge, along subsequences, as $n \to \infty$ and $\varepsilon_n \to 0$ to a minimizer of $\mathcal{E}_\infty^{(p)}$ without constraints. Of course, minimizers of $\mathcal{E}_\infty^{(p)}$ without constraints are constant functions. Hence, the labels are forgotten in the limit as $n \to \infty$. This explains why, for large n, minimizers of $\mathcal{E}_n^{(p)}$ are 'spiky'. The need to consider subsequences in the limit is due to the fact that minimizers of $\mathcal{E}_\infty^{(p)}$ without constraints are nonunique (any constant function is a minimizer).

While the degeneracy of the problem when $p \leq d$ was known, [19], 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 (the one in (16) to be precise). 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 $\varepsilon_n^p n \to \infty$ and where $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ is the empirical distribution. 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 $\{y_i\}_{i=1}^N$ is imposed through a penalty term. Our results and analysis are analogous.

Using the insights of our analysis, we define a new model which is quite similar to the original one, but for which the asymptotic consistency holds with the only upper bound requirement being that $\varepsilon_n \to 0$ as $n \to \infty$.

To prove our results we use tools from the 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 [37]. 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. We note that to the best of our knowledge, our results are the first where one proves (locally) uniform convergence of minimizers of nonlinear functionals in random discrete settings to the minimizers of the corresponding continuum functional.

Our results on the asymptotic behavior of minimizers do not provide any error estimates for finite n and do not provide precise guidance on what ε would lead to the best approximation. In Section 6, we numerically investigate

prototypical examples in one and two dimensions to shed some light on these issues. We numerically observe the predicted critical scalings for ε_n given in (4). We also numerically compare the results with our improved model (24). In investigating how precisely the observed error depends on ε we find that the error is smallest when ε is quite close to the connectivity radius on the graph. Rigorously explaining the phenomenon is, in our opinion, a valuable open problem.

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 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 and 2D numerical experiments 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 [70]. It fits in the general theme of graph-Laplacian based approaches to machine learning tasks such as clustering, which are reviewed in [65]. See also [7] for a recent application to semi-supervised learning. Zhou and Schölkopf [68] generalized the regularizers of [70] to include a version of the graph p-Laplacian. The p-Laplacian regularization has also been used by Bühler and Hein in clustering problems [9], 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 [20–22]. These papers also make the connection to the ∞-Laplacian, which is closely related to minimal Lipschitz extensions [15].

While the approach of [70] has found many applications it was pointed out by Nadler, Srebro and Zhou [48] 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 p-Laplacian 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 [19] 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 ($\varepsilon_n^p n \to 0$ as $n \to \infty$). If $p > d$ and $\varepsilon_n^p n \to \infty$ as $n \to \infty$ we prove that the problem is still degenerate 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 [69] who argue, again via regularity obtained by Sobolev embedding theorems, that taking $\alpha > \frac{d}{2}$ prevents spikes. However Dunlop, Stuart, and the authors of this paper have discovered a similar phenomenon whereby the limit may be degenerate, and spikes can occur, if ε_n converges to zero too slowly, namely if $n\varepsilon_n^{2\alpha} \to \infty$ as $n \to \infty$ [18].

García Trillos and Sanz-Alonso [35] have studied a problem in semi-supervised learning quite similar to the one in [18]. There they impose information on the data (via the observation operator) on the set of positive measure, instead of points, and obtain the convergence of posteriors in a Bayesian inverse problem on graphs with prior proportional to $\exp(-J_n(u))$, which is a richer structure than the minimizers we study. It is interesting to observe that they also require $n\varepsilon_n^{2\alpha} \to 0$ to show the convergence. Somewhat similarly, the upper bound on ε_n is used to control high frequencies. There it controls the contribution of high frequency modes of the graph Laplacian, while here we use the upper bound to establish the regularity of minimizers. For the same model as [35], García Trillos, Kaplan, Samakhoana and Sanz-Alonso [33], remove the upper bound by considering a continuum posterior defined by interpolating from the discrete space.

Our results fall in the class of asymptotic consistency results in machine learning. In general one is interested in the asymptotic behavior of an objective functional, say $E_{n,\varepsilon}(f_n)$, posed on a random sample of n points, and which also depends on a parameter ε, 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 domains $f_n : \Omega_n \to \mathbb{R}$ with those on the continuum domain $f : \Omega \to \mathbb{R}$.

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 $\lim_{\varepsilon\to 0} \lim_{n\to\infty} E_{n,\varepsilon}(f)$. Note that neither pointwise convergence or iterated pointwise convergence implies the convergence of minima/minimizers (as opposed to variational convergence). Also relevant for problems based on linear operators (which in the setting of this paper is when $p = 2$) is *spectral convergence* which asks for the eigenvalues and eigenvectors of the discrete operator to converge to eigenvalues and eigenfunctions of the continuum one. Spectral convergence is typically sufficient in linear problems for the kind of conclusions we are investigating, i.e. convergence of minima/minimizers. However for $p \neq 2$ the problems we consider here are nonlinear.

Pointwise (and similar notions of) convergence of graph Laplacians was studied by Belkin and Niyogi [6], Coifman and Lafon [14], Giné and Koltchinskii [40], Hein, Audibert and von Luxburg [42], Hein [41], Singer [55], and Ting, Huang and Jordan [62]. Spectral convergence was studied in the works of Belkin and Niyogi [6] on the convergence of Laplacian eigenmaps, von Luxburg, Belkin and Bousquet [66] and Pelletier and Pudlo [50] on graph Laplacians, and of Singer and Wu [56] 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 (i.e. it is shown that there exists a sequence $\varepsilon_n \to 0$ such that the conclusions hold as $n \to \infty$, but no bound on ϵ_n that guarantees convergence is provided). The precise and almost optimal rates were obtained in [38] 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 [10] 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 [45], and Li, Shi and Sun [46]. Error bounds for the spectral convergence of graph Laplacians have been considered by Wang [67] and García Trillos, Gerlach, Hein and one of the authors [32]. Regarding graph p-Laplacians, the authors of [19] obtain iterated pointwise convergence of graph p-Laplacians to the continuum p-Laplacian. Finally we mention that for a different type of problems, namely for nondominated sorting, Calder, Esedoğlu and Hero [13] have obtained uniform convergence of discrete solutions to the solution of a continuum Hamilton-Jacobi equation.

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 [37, 38]. In [37] 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 functionals. We refer to this paper for a description of the rich background of the works that underpin the approach. In [38] 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 [61], using a similar viewpoint. This technical setup has recently been used and extended to studies on modularity based clustering [17], Dirichlet partitions [49], Cheeger and ratio cuts [39], neighborhood graph constructions for graph cut based clustering [31], and classification problems [34, 60].

An alternative approach related to regression problems was developed by Fefferman and collaborators, Israel, Klartag and Luli, who look for a function of sufficient regularity, 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. They show that appropriate extensions exist and find efficient constructions for C^m regularity [24, 28, 29], and for Sobolev regularity [25–27]. In the context of machine learning this is a supervised learning problem and only makes use of the labeled data. In particular, the methods of

Fefferman, Israel, Klartag and Luli do not use the unlabeled data $\{x_i\}_{i=N+1}^n$.

Soon after the preprint of this paper was posted, Calder posted two preprints on closely related problems. In [11] he studied the case $p = \infty$ (also known as Lipschitz learning) and obtained consistency results which recover the problem with pointwise constraints in the large data limit. In [12] he considered semi-supervised learning with p-Laplacian regularization. He considered the game theoretic p-Laplacian which was introduced to problems on graphs in the paper by Manfredi, Oberman and Sviridov [47]. It is interesting to observe that the variational p-Laplacian (the one studied in our paper) and the game-theoretic p-Laplacian coincide in the continuum setting, but differ on graphs. While the results obtained are similar to ours, the techniques are quite different. For his results Calder used PDE based techniques, while we rely on variational techniques. He shows that the game-theoretic p-Laplacian solution of the problem has global Hölder regularity (with n independent bounds) and uses this to prove the consistency of the learning problem without an upper bound on ε_n as $n \to \infty$. He does require a more restrictive lower bound on ε_n, namely $\varepsilon_n \gg \left(\frac{\ln n}{n}\right)^{1/\max\{d+4,3d/2\}}$, for the convergence to hold.

2. Setting and Main Results. Let Ω be an open and bounded domain in $\mathbb{R}^d$. Let $\{(x_i, y_i) \,:\, i = 1, \dots, N\}$ with $x_i \in \Omega$ and $y_i \in \mathbb{R}$ be a collection of distinct labeled points. Throughout the paper we consider N to be fixed. Considering a model where N grows is an interesting problem, which we do not address here. 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 the Lebesgue measure. We assume that ρ is continuous and is bounded above and below by positive constants on Ω.

We assume that unlabeled data, $\{x_i\}_{i=N+1,\dots}$ are given by a sequence of iid samples from the 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 $\Omega_n = \{x_i\}_{i=1}^n$, 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

$$W_{ij} = \eta_\varepsilon(|x_i - x_j|) \tag{5}$$

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 $\eta(|x|) = \mathbb{I}_{|x|\le 1}$ then $\eta_\varepsilon(|x|)$ is $\frac{1}{\varepsilon^d}$ if $|x| \le \varepsilon$ 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 a penalty. We call these models *constrained* and *penalized*.

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

$$\mathcal{E}_n^{(p)}(f) = \frac{1}{\varepsilon_n^p}\frac{1}{n^2}\sum_{i,j=1}^n W_{ij}|f(x_i) - f(x_j)|^p \tag{6}$$

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

For technical reasons it is convenient to include the constraints in the functional $\mathcal{E}_n^{(p)}$ and hence we define

$$\mathcal{E}_{n,\mathrm{con}}^{(p)}(f) = \begin{cases} \frac{1}{\varepsilon_n^p}\frac{1}{n^2}\sum_{i,j=1}^n W_{ij}|f(x_i) - f(x_j)|^p & \text{if } f(x_i) = y_i \text{ for } i = 1, 2, \dots, N \\ \infty & \text{else.} \end{cases} \tag{7}$$

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

$$R^{(q)}(f) = \sum_{i=1}^N |y_i - f(x_i)|^q. \tag{8}$$

We define the penalized estimator as the minimizer of

$$\mathcal{S}_n^{(p)}(f) = \mathcal{E}_n^{(p)}(f) + \lambda R^{(q)}(f) \tag{9}$$

over all functions $f : \Omega_n \to \mathbb{R}$, where $\lambda > 0$ is a tunable parameter.

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

$$\mathcal{E}_{\infty}^{(p)}(f) = \begin{cases} \sigma_\eta \int_\Omega |\nabla f(x)|^p \, \rho^2(x) \, \mathrm{d}x & \text{if } f \in W^{1,p}(\Omega), \\ \infty & \text{else.} \end{cases} \tag{10}$$

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

$$\mathcal{E}_{\infty,\mathrm{con}}^{(p)}(f) = \begin{cases} \mathcal{E}_{\infty}^{(p)}(f) & \text{if } f \in W^{1,p}(\Omega) \text{ and } f(x_i) = y_i \text{ for } i = 1, \dots, N \\ \infty & \text{else.} \end{cases} \tag{11}$$

The constant σ_η above is defined, using $e_1 = [1, 0, \dots, 0]^T$, by

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

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

$$\mathcal{S}_{\infty}^{(p)}(f) = \mathcal{E}_{\infty}^{(p)}(f) + \lambda R^{(q)}(f). \tag{12}$$

The functional $\mathcal{S}_{\infty}^{(p)}$ is well defined whenever $p > d$.

We note that functionals (11) and (12) are lower semi-continuous with respect to the L^p norm. In addition, coercivity of both functionals follows from Sobolev embeddings. Coercivity and lower semi-continuity imply existence of minimizers, e.g. [30, Theorem 3.6]. Strict convexity implies that the minimizers are unique.

We are interested in asymptotic behavior of minimizers f_n of the discrete models, say $\mathcal{E}_{n,\mathrm{con}}^{(p)}$. We say that $\mathcal{E}_{n,\mathrm{con}}^{(p)}$ is *asymptotically consistent* with $\mathcal{E}_{\infty,\mathrm{con}}^{(p)}$ if the minimizers f_n of $\mathcal{E}_{n,\mathrm{con}}^{(p)}$ converge as $n \to \infty$ to a minimizer of $\mathcal{E}_{\infty,\mathrm{con}}^{(p)}$. One should note the 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 [37] 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 [51] is $\varepsilon_n \sim \left(\frac{\ln n}{n}\right)^{\frac{1}{d}}$. We require (when $d \geq 3$) $\varepsilon_n \gg \left(\frac{\ln n}{n}\right)^{\frac{1}{d}}$ which means that our lower bound is almost optimal. We have discovered that if $\varepsilon_n \to 0$ too slowly then the discrete functional $\mathcal{E}_{n,\mathrm{con}}^{(p)}$ lacks sufficient regularity for the constraints to be preserved in the limit. The optimal upper bound on ε_n is discussed after 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 and $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ is the empirical measure;
- **(A5)** ε_n is a sequence converging to 0 satisfying the lower bound

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

- **(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 $\sigma_\eta = \int_{\mathbb{R}^d} \eta(|w|)|w \cdot e_1|^p \, \mathrm{d}w < \infty$).

We note that Assumption **(A2)** implies that μ is equivalent to the Lebesgue measure on Ω (denoted $\mathcal{L}\lfloor_\Omega$). Hence, $f \in L^p(\Omega)$ if and only if $f \in L^p(\mu)$, where with an abuse of notation we write $L^p(\Omega) := L^p(\mathcal{L}\lfloor_\Omega)$.

The first main result of the paper is the following theorem. The 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 Assumptions **(A1) - (A8)**. Let f_n be a sequence of minimizers of $\mathcal{E}^{(p)}_{n,\mathrm{con}}$ defined in (7) with graph weights W_{ij} given by (5). 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) the whole sequence f_n converges to f both in TL^p and locally uniformly, meaning that for any Ω' with $\overline{\Omega'} \subset \Omega$

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

(b) f is a minimizer of $\mathcal{E}^{(p)}_{\infty,\mathrm{con}}$ defined in (11);

(ii) if $n\varepsilon_n^p \to \infty$ as $n \to \infty$ then f is a minimizer of $\mathcal{E}^{(p)}_{\infty}$ defined in (10).

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 continuous, and thus it is possible to impose pointwise values of f, as needed to define $\mathcal{E}^{(p)}_{\infty,\mathrm{con}}$ in (11).

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. This is due to f_n developing narrow spikes near labeled points $\{x_i\}_{i=1}^N$ and becoming nearly constant everywhere else. In the TL^p limit the spikes disappear.

This motivates referring to the scaling $n\varepsilon_n^p \to \infty$ 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 $p > 1$ and $q > 0$. Assume Ω, μ, η, and x_i satisfy Assumptions **(A1) - (A8)**. Let f_n be a sequence of minimizers of $\mathcal{S}^{(p)}_{n}$ defined in (9) where $\lambda > 0$, $R^{(q)}$ is given by (8), and $\mathcal{E}^{(p)}_{n}$ is given by (6) with graph weights W_{ij} given by (5). 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) the whole sequence f_n converges to f both in TL^p and locally uniformly, meaning that for any Ω' with $\overline{\Omega'} \subset \Omega$

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

(b) f is a minimizer of $\mathcal{S}^{(p)}_{\infty}$ defined in (12);

(ii) if $n\varepsilon_n^p \to \infty$ as $n \to \infty$ then f is a minimizer of $\mathcal{E}^{(p)}_{\infty}$ defined in (10).

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 $\mathcal{E}^{(1)}_{\infty}$ would be a weighted TV semi-norm $\mathcal{E}^{(1)}_{\infty} = \sigma_\eta TV(\cdot;\rho)$ 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 [37].

In Theorem 2.1 we proved that the model $\mathcal{E}^{(p)}_{n,\mathrm{con}}$, defined in (7), is consistent as $n \to \infty$ and lower bounds **(A5)** hold, only if

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

This upper bound is undesirable as it restricts the range of ε that can be used. In our numerical experiments (see Figures 2(a) and 3(a)) we observe that the range of ε for which the limiting problem is well approximated can be

quite narrow. This problem is particularly pronounced if $p > d$ is close to d, which is the regime identified in [19] as the most relevant for semi-supervised learning.

This motivated us to explore changing the model in such a way that it remains asymptotically consistent with $\mathcal{E}^{(p)}_{\infty,\mathrm{con}}$, but does not require an upper bound on ε_n (other than $\varepsilon_n \to 0$ as $n \to \infty$). In Section 5 we introduce a new, related, model which has the desired properties. Furthermore its minimizers can be computed with almost identical algorithms as those for $\mathcal{E}^{(p)}_{n,\mathrm{con}}$. More precisely, the minimization procedure remains unchanged, and one only has to increase the size of the constraint set. The new model is asymptotically consistent as $n \to \infty$, and $\varepsilon_n \to 0$, whenever **(A5)** holds.

While in the original model the minimizers of the objective functional for a fixed, large n have a desired behavior in a sometimes narrow band of admissible ε_n, we observe in numerical experiments (see for example Figure 5) that for the improved model the error grows gradually with ε_n, as opposed to becoming catastrophic as ε_n reaches a certain threshold.

3. Background Material. In an effort to make this paper more self-contained we briefly review three key ideas 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 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 [8, 16] 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, $L^0(Z; \mathbb{R} \cup \{\pm\infty\})$ be the set of measurable functions from Z to $\mathbb{R} \cup \{\pm\infty\}$, and $(\mathcal{X}, \mathbb{P})$ be a probability space. The function $\mathcal{X} \ni \omega \mapsto E^{(\omega)}_n \in L^0(Z; \mathbb{R} \cup \{\pm\infty\})$ is a random variable. We say $E^{(\omega)}_n$ Γ-converge almost surely on the domain Z to $E_\infty : Z \to \mathbb{R} \cup \{\pm\infty\}$ with respect to d, and write $E_\infty = \Gamma\text{-}\lim_{n\to\infty} E^{(\omega)}_n$, 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

$$E_\infty(f) \le \liminf_{n\to\infty} E^{(\omega)}_n(f_n), \text{ and}$$

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

$$E_\infty(f) \ge \limsup_{n\to\infty} E^{(\omega)}_n(f_n).$$

For ease of notation we will suppress the dependence of ω on on our functionals, that is we apply the above definition to $E_n = \mathcal{E}^{(p)}_n$. The almost sure statement in the above definition does not play a significant role in the proofs. Essentially it is enough to consider the set of realizations 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 it implies the convergence of minimizers. The proof can be found in [8, Theorem 1.21] or [16, 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 ∞ 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. Section 3.3 overviews the metric space we use to analyze the asymptotics of our semi-supervised learning models, in particular it allows us to go from discrete to continuum.

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 [64] and Santambrogio [53].

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

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

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,

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

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

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

Optimal transportation plays an important role in comparing the discrete and continuum objects we study. In particular, we use sharp estimates on the ∞-optimal transportation distance between a measure and the empirical measure of its sample. In the form below, for $d \ge 2$, they were established in [36], which extended the related results in [1, 43, 54, 57]. For $d = 1$ the estimates are simpler, and follow from the law of iterated logarithms. More precisely the map $T_n(x) = x_i$, where $x \in \left(z_{i-1}^{(n)}, z_i^{(n)}\right]$ and $\mu((-\infty, z_i^{(n)}]) = i/n$, is the optimal transport map (assuming x_i are ordered: $x_i \le x_{i+1}$), and one can show $\inf_{x\in\Omega} \rho(x) \|T_n - \mathrm{Id}\|_{L^\infty(\Omega)} \le \|F - \hat{F}_n\|_{L^\infty(\Omega)}$ where $F, \hat{F}_n$ are the cumulative distribution functions for μ, μ_n. The empirical cumulative distribution function converges in L^∞ with rate $\sqrt{\frac{\ln\ln n}{n}}$, see [63, Section 19.1] by 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 $x_1, x_2, \ldots$ be a sequence of independent random variables with distribution μ and let μ_n be the empirical measure. Then, there exists 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 with the property*

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

where

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

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 $\Omega_n = \{x_i\}_{i=1}^n$, 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 from which the data points are sampled. One would like to compare f and f_n in a way that is consistent with the L^p topology. To do so we use the TL^p space that was introduced in [37], 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 [38, 58, 59].

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

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

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, [37].

As shown in [37], 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,

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

where the push forward of the measure $T_{\#}\mu$ is defined in Section 3.2. This formula provides an ~~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 the $TL^p(\overline{\Omega})$ space is not complete, and that its completion was discussed in [37]. 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 [38]) is that constraints in $\mathcal{E}^{(p)}_\infty$ are imposed pointwise on a set of μ measure zero. [The reason that these constraints make sense is that for $p > d$ the finiteness of $\mathcal{E}^{(p)}_\infty(f)$ implies that f is continuous.] We note that the TL^p convergence used in [38] 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, that is how to show that uniform boundedness of $\mathcal{E}^{(p)}_{n,\mathrm{con}}(f_n)$ implies the existence of a (locally) uniformly converging subsequence. Our approach combines discrete and continuum regularity results. Namely, we obtain in Lemma 4.1 a local control of oscillations of f_n over distances of order ε_n. In Lemma 4.2 we show that discrete functionals $\mathcal{E}^{(p)}_n(f_n)$ control the values of the associated nonlocal continuum functionals $\mathcal{E}^{(NL,p)}_{\varepsilon_n}(\tilde{f}_n)$ (defined in (16) 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, namely for a fixed $f \in C^1(\Omega)$, $\mathcal{E}^{(p)}_n(f)$ is the Monte Carlo integral approximation of $\mathcal{E}^{(NL,p)}_{\varepsilon_n}(f)$ which approximates $\mathcal{E}^{(p)}_\infty(f)$ only as $\varepsilon_n \to 0$. In particular, the issue is that the nonlocal functionals do not share the regularizing properties of the limiting functional. Only a weaker form of regularity holds. ~~In particular,~~ We show in Lemma 4.3 that control of the nonlocal energy is sufficient to provide regularity at scales larger than ε_n. Combining these estimates is enough to imply the compactness with respect to (locally) uniform convergence in Lemma 4.5.

We state the discrete regularity results in a deterministic setting for finite, fixed n, under the assumption on the smallness of $d_\infty(\mu, \mu_n)$. When applying the lemma in the random setting this condition will hold due to Theorem 3.3.

LEMMA 4.1 (discrete regularity). *Let $p > 1$. Suppose that Ω and μ satisfy Assumptions **(A1) - (A2)** and η satisfies Assumptions **(A6) - (A8)**. Let $\varepsilon_n > 0$ and $x_i \in \Omega$ for $i = 1, \dots, n$. Let $\mathcal{E}^{(p)}_n$ be defined by (6), with graph weights W_{ij} given by (5), and $\Omega_n = \{x_i\}_{i=1}^n$. For $f_n : \Omega_n \to \mathbb{R}$, we define $\mathrm{osc}^{(n)}_\varepsilon(f_n) : \Omega_n \to \mathbb{R}$ by*

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

Then, for all $\alpha_0 > 0$, there exists $b > 0$ (depending only on η and α_0) and $C > 0$ such that whenever $d_\infty(\mu, \mu_n) \le \frac{1}{8} b\varepsilon_n$:

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

for all $\alpha \ge \alpha_0$, all $f_n : \Omega_n \to \mathbb{R}$, and all $k \in \{1, 2, \dots, n\}$

Proof. Let $\tilde{\eta} : [0,\infty) \to [0,\infty)$ be defined by $\tilde{\eta}(t) = a$ if $0 \le t < b$ and $\tilde{\eta}(t) = 0$ otherwise, where a and b are chosen such that $\tilde{\eta} \le \eta$. We can furthermore choose b so that $b \le \alpha_0$. For all $k \in \{1,\dots,n\}$ 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 $\operatorname{osc}^{(n)}_{\frac{b\varepsilon_n}{2}}(f_n)(x_k) = \bar{f}_n(x_k) - \underline{f}_n(x_k)$ and for all $x \in B\left(x_k, \frac{b\varepsilon_n}{2}\right) \cap \Omega_n$

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

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,

$$\begin{aligned}
\mathcal{E}^{(p)}_n(f_n) &\ge \frac{1}{\varepsilon_n^{p+d} n^2} \sum_{i,j=1}^n \tilde{\eta}\left(\frac{|x_i - x_j|}{\varepsilon_n}\right) |f_n(x_i) - f_n(x_j)|^p \\
&\ge \frac{1}{\varepsilon_n^{p+d} n^2} \sum_{j : |x_j - \bar{x}_k| \le b\varepsilon_n} \tilde{\eta}\left(\frac{|\bar{x}_k - x_j|}{\varepsilon_n}\right) |f_n(x_j) - f_n(\bar{x}_k)|^p \\
&\ge \frac{a}{\varepsilon_n^{p+d} n^2} \sum_{j : |x_j - x_k| \le \frac{b\varepsilon_n}{2}} |f_n(x_j) - f_n(\bar{x}_k)|^p, \qquad \text{since } |x_k - \bar{x}_k| \le \frac{b\varepsilon_n}{2} \\
&\ge \frac{a}{2^{p+1}\varepsilon_n^{p+d} n^2} \left(\operatorname{osc}_{\frac{b\varepsilon_n}{2}}(f_n)(x_k)\right)^p \# \left\{ j \,:\, |x_j - x_k| \le \frac{b\varepsilon_n}{2} \right\} \\
&= \frac{a}{2^{p+1}\varepsilon_n^{p+d} n} \left(\operatorname{osc}_{\frac{b\varepsilon_n}{2}}(f_n)(x_k)\right)^p \mu_n\left(B\left(x_k, \frac{b\varepsilon_n}{2}\right)\right).
\end{aligned} \tag{13}$$

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

$$\begin{aligned}
\frac{1}{\varepsilon_n^d}\mu_n\left(B\left(x_k, \frac{b\varepsilon_n}{2}\right)\right) &= \frac{1}{\varepsilon_n^d}\int_\Omega \mathbb{I}_{\{|T_n(x) - x_k| \le \frac{b\varepsilon_n}{2}\}} \rho(x)\,\mathrm{d}x \\
&\ge \frac{\inf_{x\in\Omega}\rho(x)}{\varepsilon_n^d}\int_\Omega \mathbb{I}_{\{|x - x_k| \le \frac{b\varepsilon_n}{2} - \|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}\}}\,\mathrm{d}x \\
&= \left(\inf_{x\in\Omega}\rho(x)\right)\mathrm{Vol}\left(B\left(0, \frac{b}{2} - \frac{\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}}{\varepsilon_n}\right)\right).
\end{aligned} \tag{14}$$

Since $d_\infty(\mu,\mu_n) \le \frac{1}{8}b\varepsilon_n$, combining (13) and (14) gives

$$\left(\operatorname{osc}_{\frac{b\varepsilon_n}{2}}(f_n)(x_k)\right)^p \le \frac{2^{p+1}\varepsilon_n^p n \mathcal{E}^{(p)}_n(f_n)}{a\left(\inf_{x\in\Omega}\rho(x)\right)\mathrm{Vol}\left(B\left(0, \frac{3b}{8}\right)\right)} =: C_1 \varepsilon_n^p n \mathcal{E}^{(p)}_n(f_n). \tag{15}$$

Note that the constant C_1 does not depend on n, ε_n or x_k.

Let $\gamma = \frac{b\varepsilon_n}{2}$ and let $m = \lceil \frac{8\alpha}{b} \rceil$. Let

$$x^* \in \operatorname*{argmax}_{z \in B(x_k, \alpha\varepsilon_n) \cap \Omega_n} f_n(z) \quad \text{and} \quad x_* \in \operatorname*{argmin}_{z \in B(x_k, \alpha\varepsilon_n) \cap \Omega_n} f_n(z).$$

For $j = 0,\dots,m$ let $z_j = \frac{m-j}{m}x_* + \frac{j}{m}x^*$. Note that $|z_{j+1} - z_j| < 2\varepsilon_n\alpha/m \le \gamma/2$. Since $d_\infty(\mu,\mu_n) \le \frac{1}{4}b\varepsilon_n$, for all $j = 0,\dots,m$ there exists $x_{i_j} \in \Omega_n \cap B(z_j, \frac{\gamma}{2})$. We observe that $|x_{i_{j-1}} - z_j| \le |x_{i_{j-1}} - z_{j-1}| + |z_{j-1} - z_j| < \gamma$.

Therefore,

$$\begin{aligned}\operatorname{osc}_{\alpha\varepsilon_n}(f_n)(x_k) &= |f(x^*) - f(x_*)| \\ &\le \sum_{j=1}^{m} |f(x_{i_j}) - f(x_{i_{j-1}})| \\ &\le \sum_{j=1}^{m} \operatorname{osc}_{\gamma}(f_n)(z_j) \\ &\le \left\lceil \frac{8\alpha}{b} \right\rceil \sup_{x_i \in \Omega_n} \operatorname{osc}_{\gamma}(f_n)(x_i).\end{aligned}$$

Using that $\alpha_0 \ge b$, for $\alpha > \alpha_0$ from (15) it follows that

$$(\operatorname{osc}_{\alpha\varepsilon_n}(f_n)(x_k))^p \le \left(\frac{9\alpha}{b}\right)^p \left(\sup_{x_i} \operatorname{osc}_{\frac{b\varepsilon_n}{2}}(f_n)(x_i)\right)^p \le C\alpha^p \varepsilon_n^p n \mathcal{E}_n^{(p)}(f_n)$$

where $C = C_1 \left(\frac{9}{b}\right)^p$. ∎

LEMMA 4.2 (discrete to nonlocal control). *Let $p \ge 1$. Assume Ω, μ, η, and x_i satisfy Assumptions **(A1) - (A8)**. Let constants $a, b > 0$ be such that $\tilde{\eta}(|x|) = a$ for $|x| < b$ and $\tilde{\eta}(|x|) = 0$ otherwise, and so that $\tilde{\eta} \le \eta$. Let T_n be a sequence of transport maps satisfying the conclusions of Theorem 3.3 and let $\tilde{\varepsilon}_n = \varepsilon_n - \frac{2\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}}{b}$. Define $\mathcal{E}_n^{(p)}(\cdot;\eta)$ by (6), with graph weights W_{ij} given by (5), and where we explicitly denote the dependence of η. 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

$$\mathcal{E}_{\varepsilon}^{(NL,p)}(f;\eta) = \frac{1}{\varepsilon^p} \int_\Omega \int_\Omega \eta_\varepsilon(|x-z|)|f(x) - f(z)|^p \rho(x)\rho(z)\,\mathrm{d}x\,\mathrm{d}z. \tag{16}$$

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

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

So,

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

and therefore

$$\left|\frac{x-z}{\tilde{\varepsilon}_n}\right| < b \quad \Rightarrow \quad \tilde{\eta}\left(\frac{|x-z|}{\tilde{\varepsilon}_n}\right) = a = \tilde{\eta}\left(\frac{|T_n(x) - T_n(z)|}{\varepsilon_n}\right).$$

Hence,

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

Now,

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

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

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

LEMMA 4.3 (nonlocal to averaged local). *Let $p \geq 1$ and assume that Ω and μ satisfy Assumptions **(A1) - (A2)**, and η satisfies Assumptions **(A6) - (A8)**. Then, there exists a constant $C \geq 1$ and a radially symmetric mollifier J with $\operatorname{supp}(J) \subseteq \overline{B(0,1)}$ such that for all $\varepsilon > 0$, $f \in L^p(\Omega)$, and any $\Omega' \subset\subset \Omega$ (i.e. for every Ω' that is compactly contained in Ω) with $\operatorname{dist}(\Omega', \partial\Omega) > \varepsilon$ it holds that*

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

where $\mathcal{E}_\infty^{(p)}$ is defined by (10) and $\mathcal{E}_\varepsilon^{(NL,p)}$ is defined by (16), and for both functionals we explicitly denote the dependence on the domain (rather than the kernel η).

Proof. By the continuity of η at zero we can find some $b \in (0,1]$ such that $\eta(|x|) > \eta(0)/2 > 0$ for all $|x| \leq b$. Let J be a radially symmetric mollifier whose support is contained in $\overline{B(0,b)}$. There exists $\beta > 0$ such that $J \leq \beta\eta(|\cdot|)$ and $|\nabla J| \leq \beta\eta(|\cdot|)$. Without loss of generality we can assume $\operatorname{supp}(\eta) \subset [0,1]$. 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, for $\gamma_\eta = \int_{B(0,1)} \eta(|w|)\,\mathrm{d}w$,

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

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

$$\begin{aligned}
\int_{\Omega'} |\nabla g_\varepsilon(x)|^p\, \rho^2(x)\,\mathrm{d}x &\leq \gamma_\eta^{p-1}\beta^p \int_\Omega \int_\Omega \eta_\varepsilon(|x-z|) \left|\frac{f(z) - f(x)}{\varepsilon}\right|^p \rho^2(x)\,\mathrm{d}z\,\mathrm{d}x \\
&\leq \frac{\gamma_\eta^{p-1}\beta^p \sup_{x\in\Omega} \rho(x)}{\inf_{x\in\Omega} \rho(x)} \mathcal{E}_\varepsilon^{(NL,p)}(f; \Omega)
\end{aligned}$$

which completes the proof. ∎

We now establish the compactness property for sequences bounded in L^∞. The result follows by combining Lemma 4.2, known results in the literature and a simple interpolation argument.

PROPOSITION 4.4 (compactness). *Let $p > 1$. Assume Ω, μ, η, and x_i satisfy Assumptions **(A1) - (A8)**. Let $\mathcal{E}_n^{(p)}$ be defined by (6), with graph weights W_{ij} given by (5), and $\Omega_n = \{x_i\}_{i=1}^n$. 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 $f \in L^p(\mu)$.*

Restricting the space to the set of functions bounded in L^∞ is needed since the functionals $\mathcal{E}_n^{(p)}$ are invariant under adding a constant. When applying this proposition to prove Theorem 2.1, this restriction is not an issue since both discrete and continuum minimizers of the constrained functionals satisfy an L^∞ bound. Nevertheless let us briefly discuss the compactness of the functionals with constraints, $\mathcal{E}_{n,\mathrm{con}}^{(p)}(f_n)$. If $\lim_{n\to\infty} n\varepsilon_n^p = \infty$ (i.e. in the degenerate case) assuming an L^∞ bound is essential since the limiting functional $\mathcal{E}_\infty^{(p)}$ is invariant under adding a constant and thus the loss of constraints in the limit which occurs when $\lim_{n\to\infty} n\varepsilon_n^p = \infty$ would lead to loss of compactness. When $\lim_{n\to\infty} n\varepsilon_n^p = 0$ then the constraints are present in the limit and in particular there is enough regularity to infer a bound on $\sup_{n\in\mathbb{N}} \|f_n\|_{L^\infty(\mu_n)}$ whenever $\sup_{n\in\mathbb{N}} \mathcal{E}_{n,\mathrm{con}}^{(p)}(f_n) < \infty$. Hence, when $\lim_{n\to\infty} n\varepsilon_n^p = 0$ one could remove the assumption on L^∞ bounds in a compactness result for $\mathcal{E}_{n,\mathrm{con}}^{(p)}$.

Proof of Proposition 4.4. As in Lemma 4.2, let T_n be a sequence of transport maps satisfying the conclusions of Theorem 3.3 and let $\tilde{\varepsilon}_n = \varepsilon_n - \frac{2\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}}{b}$. Consider also $\tilde{\eta}$ of Lemma 4.2. Then $\mathcal{E}_{\tilde{\varepsilon}_n}^{(NL,p)}(f_n \circ T_n; \tilde{\eta}) \leq C\mathcal{E}_n^{(p)}(f_n; \eta)$. By Jensen's inequality, $\mathcal{E}_{\tilde{\varepsilon}_n}^{(NL,p)}(f_n \circ T_n; \tilde{\eta}) \geq c\left(\mathcal{E}_{\tilde{\varepsilon}_n}^{(NL,1)}(f_n \circ T_n; \tilde{\eta})\right)^p$ for some $c > 0$ and all n. By the proof of Theorem 1.2 in [37], $f_n \circ T_n$ is precompact in $L^1(\mu)$ and thus (μ_n, f_n) is precompact in TL^1.

We note that from the proof of Theorem 1.2 in [37] it follows that there in fact exists a subsequence f_{n_m}, and a sequence of transportation maps $(T_{n_m})_\# \mu = \mu_{n_m}$ 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 $\|f - f_{n_m} \circ T_{n_m}\|_{L^\infty(\mu)} \leq \|f\|_{L^\infty(\mu)} + \sup_{n\in\mathbb{N}} \|f_n\|_{L^\infty(\mu_n)} < \infty$, 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 Proposition 4.4. Assume that $\varepsilon_n^p n \to 0$ as $n \to \infty$ (which, due to **(A5)** implies that $p > d$), $(\mu_n, f_n) \to (\mu, f)$ in the TL^p metric as $n \to \infty$, and $\sup_{n\in\mathbb{N}} \mathcal{E}_n^{(p)}(f_n) < \infty$. Then $f \in C^{0,\gamma}(\Omega)$, with $\gamma = 1 - \frac{d}{p} > 0$, and for all Ω' compactly supported in Ω*

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

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. Consider constants $a, b > 0$ such that $\tilde{\eta}(t) := a$ if $|t| \leq b$ and $\tilde{\eta}(t) := 0$ if $|t| > b$ satisfies $\tilde{\eta} \leq \eta$. We define $\tilde{f}_n = f_n \circ T_n$ where T_n is the transportation map satisfying the conclusions of Theorem 3.3 and set $\tilde{\varepsilon}_n = \varepsilon_n - \frac{2\|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}}{b}$. Then, for n sufficiently large $\tilde{\varepsilon}_n > 0$, and $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$. Let $\mathcal{E}_{\tilde{\varepsilon}_n}^{(NL,p)}$ be the non-local Dirichlet energy defined in (16). Then, by Lemma 4.2

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

Hence, $\mathcal{E}_{\tilde{\varepsilon}_n}^{(NL,p)}(\tilde{f}_n; \tilde{\eta})$ is bounded. Therefore, by Lemma 4.3, we have that $\mathcal{E}_\infty^{(p)}(J_{\tilde{\varepsilon}_n} * \tilde{f}_n; \Omega')$ (where J is a mollifier with the properties given in Lemma 4.3) is bounded for every $\Omega' \subset\subset \Omega$ (i.e. ~~for every~~ Ω' that is compactly contained in Ω). Furthermore $\|J_{\tilde{\varepsilon}_n} * \tilde{f}_n\|_{L^p(\Omega')} \leq \|\tilde{f}_n\|_{L^p(\Omega)}$ (see for example [44, Theorem C.19(iii)]). From this L^p bound and

$$\mathcal{E}_\infty^{(p)}(J_{\tilde{\varepsilon}_n} * \tilde{f}_n; \tilde{\eta}, \Omega') \geq \sigma_\eta \left(\inf_{x\in\Omega} \rho^2(x)\right) \left\|\nabla(J_{\tilde{\varepsilon}_n} * \tilde{f}_n)\right\|_{L^p(\Omega')}$$

it follows that $J_{\tilde{\varepsilon}_n} * \tilde{f}_n$ is locally bounded in $W^{1,p}$, i.e. $\sup_{n\in\mathbb{N}} \|J_{\tilde{\varepsilon}_n} * \tilde{f}_n\|_{W^{1,p}(\Omega')} < \infty$. 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 $C^{0,\gamma}$ for $\gamma = 1 - \frac{d}{p}$ (Morrey's inequality, for example see Theorem 4 in Section 5.6.2 of [23]), 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 $B(x_k, \delta)$ for any δ such that $B(x_k, \delta) \subset\subset \Omega$. Condition **(A5)** and Theorem 3.3 imply that with probability one, for all n large enough the requirement $d_\infty(\mu, \mu_n) \leq \frac{1}{8} b \varepsilon_n$ of Lemma 4.1 is satisfied. Thus, for all n sufficiently large, for all $x \in B(x_k, 2\tilde{\varepsilon}_n) \cap \Omega_n$ we have

$$|f_n(x_k) - f_n(x)| \leq \operatorname{osc}_{2\tilde{\varepsilon}_n}(f_n)(x_k) \leq \operatorname{osc}_{3\varepsilon_n}(f_n)(x_k) \leq \left(C \mathcal{E}_n^{(p)}(f_n) n \varepsilon_n^p \right)^{\frac{1}{p}}$$

where $C > 0$ is independent of n. It follows that

$$\max_{\{k \,:\, x_k \in \Omega'\}} \max_{x \in B(x_k, 2\tilde{\varepsilon}_n) \cap \Omega_n} |f_n(x) - f_n(x_k)| \to 0.$$

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

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

and the above converges to zero.

Clearly, if $f_n(x_i) = y_i$ for all n then, choosing Ω' sufficiently large such that $x_i \in \Omega'$, we have

$$|f(x_i) - y_i| = |f(x_i) - f_n(x_i)| \leq \max_{\{k \,:\, x_k \in \Omega'\}} |f(x_k) - f_n(x_k)| \to 0.$$

Hence, $f(x_i) = y_i$. ∎

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 [37]. 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 $\mathcal{E}_{n,\mathrm{con}}^{(p)}$ towards $\mathcal{E}_\infty^{(p)}$ or $\mathcal{E}_{\infty,\mathrm{con}}^{(p)}$ depending on how quickly $\varepsilon_n \to 0$ as $n \to \infty$. Steps in the proof of this claim rely on Lemma 4.6.

LEMMA 4.6 (continuum nonlocal to local). *Let $p > 1$. Suppose that Ω and μ satisfy Assumptions **(A1) - (A2)** and η satisfies Assumptions **(A6) - (A8)**. Then $\mathcal{E}_\varepsilon^{(NL,p)}$, defined in (16), Γ-converges as $n \to \infty$ in $L^p(\Omega)$ to the functional $\mathcal{E}_\infty^{(p)}$ defined in (10).*

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 [52]. We remark that while the functional in [52] 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 [37] and has been remarked in Proposition 1.10 in [38].

We now state the Γ-convergence result. Its proof is divided into two lemmas below. The corresponding compactness property has already been established in Proposition 4.4.

THEOREM 4.7 (discrete to local Γ-convergence). *Let $p > 1$. Suppose that Ω, μ, η, ε_n, and x_i satisfy Assumptions **(A1) - (A8)**. Let $M \geq \max_{i=1,\dots,N} |y_i|$. Then, with probability one, $\mathcal{E}_{n,\mathrm{con}}^{(p)}$, defined in (7) with graph weights W_{ij} given by (5), Γ-converges as $n \to \infty$ in the TL^p metric on the set $\{(\nu, g) \,:\, \nu \in \mathcal{P}(\Omega), \|g\|_{L^\infty(\nu)} \leq M\}$ to the functional*

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

where $\mathcal{E}_\infty^{(p)}$ is defined by (10) and $\mathcal{E}_{\infty,\mathrm{con}}^{(p)}$ is defined by (11).

We prove the liminf inequalities and the existence of a recovery sequence separately. Since $\mathcal{E}_\infty^{(p)} \le \mathcal{E}_{\infty,\mathrm{con}}^{(p)}$ 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 $f \in L^p(\mu)$ 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*

$$\mathcal{E}_\infty^{(p)}(f) \le \liminf_{n\to\infty} \mathcal{E}_n^{(p)}(f_n) \le \liminf_{n\to\infty} \mathcal{E}_{n,\mathrm{con}}^{(p)}(f_n). \tag{17}$$

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

$$\mathcal{E}_{\infty,\mathrm{con}}^{(p)}(f) \le \liminf_{n\to\infty} \mathcal{E}_{n,\mathrm{con}}^{(p)}(f_n). \tag{18}$$

Proof. Let $f_n \to f$ in TL^p. The first inequality of (17) follows from Lemma 4.6 in the same way the analogous result is shown for $p = 1$ in Section 5 of [37]. The second inequality follows from the definition of $\mathcal{E}_n^{(p)}$ and $\mathcal{E}_{n,\mathrm{con}}^{(p)}$.

When $\lim_{n\to\infty} n\varepsilon_n^p = 0$ the inequality (18) is a consequence of Lemma 4.5. ∎

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

LEMMA 4.9. *Under the same conditions as Theorem 4.7, with probability one, for any function $f \in L^p(\mu)$, 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*

$$\mathcal{E}_{\infty,\mathrm{con}}^{(p)}(f) \ge \limsup_{n\to\infty} \mathcal{E}_{n,\mathrm{con}}^{(p)}(f_n). \tag{19}$$

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

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

Proof. The proof of the first inequality is a straightforward adaptation of the analogous result for $p = 1$ in Section 5 of [37]. 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 $f \in L^p(\mu)$, 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

$$\mathcal{E}_\infty^{(p)}(f) \ge \limsup_{n\to\infty} \mathcal{E}_n^{(p)}(f_n). \tag{21}$$

Let us now consider the case that $n\varepsilon_n^p \to \infty$ as $n \to \infty$ and show the second inequality. Suppose $\mathcal{E}_\infty^{(p)}(f) < \infty$ else the lemma is trivial. Let f_n be the recovery sequence for (21).

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 $\|\hat{f}_n\|_{L^\infty(\mu_n)} \le M$. To show (20) it suffices to show that

$$\lim_{n\to\infty} \left(\mathcal{E}_n^{(p)}(f_n) - \mathcal{E}_{n,\mathrm{con}}^{(p)}(\hat{f}_n) \right) = 0. \tag{22}$$

We may write,

$$\begin{aligned} \left| \mathcal{E}_n^{(p)}(f_n) - \mathcal{E}_{n,\mathrm{con}}^{(p)}(\hat{f}_n) \right| &\le \frac{1}{\varepsilon_n^p} \frac{2}{n^2} \sum_{i=1}^{N} \sum_{j=1}^{n} \eta_{\varepsilon_n}(|x_i - x_j|) \, \big| \, |f(x_i) - f(x_j)|^p - |y_i - f(x_j)|^p \big| \\ &\le \frac{2^{p+1} M^p}{\varepsilon_n^p n} \sum_{i=1}^{N} \frac{1}{n} \sum_{j=1}^{n} \eta_{\varepsilon_n}(|x_i - x_j|) \end{aligned} \tag{23}$$

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

$$\mu_n(B(x_i, \varepsilon_n b)) = \mu\left(\{x \,:\, |T_n(x) - x_i| < \varepsilon_n b\}\right) \leq \mu\left(\left\{x \,:\, |x - x_i| < \varepsilon_n b + \|\mathrm{Id} - T_n\|_{L^\infty(\Omega)}\right\}\right),$$

then, using Theorem 3.3,

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

Combining this inequality with (23) implies (22).
Step 2. Now consider 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 $\|\mathrm{Id} - T_n\|_{L^\infty(\Omega)} \leq \varepsilon_n$. Consequently,

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

where the penultimate inequality follows from

$$\frac{|x_i - T_n(y)|}{\varepsilon_n} \geq \left(\frac{|x_i - y| - \|T_n - \mathrm{Id}\|_{L^\infty(\Omega)}}{\varepsilon_n}\right)_+ \geq \left(\frac{|x_i - y|}{\varepsilon_n} - 1\right)_+.$$

Again combining this estimate with (23) implies (22). ∎

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 $f \in W^{1,p}$ with $f_{n_m} \to f$ 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 $\|f_n\|_{L^\infty(\mu_n)} \leq M$,*

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

where the Γ-convergence is considered in the 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 $\|f_n\|_{L^\infty(\mu_n)} \leq M$,*

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

where the Γ-convergence is considered in the 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 (17) of Lemma 4.8 we have $\liminf_{n\to\infty} \left(\mathcal{E}_n^{(p)}(f_n) + \lambda R^{(q)}(f_n)\right) \geq \mathcal{E}_\infty^{(p)}(f) + \lambda R^{(q)}(f)$. The limsup inequality follows in a similar manner from equation (21) and Lemma 4.5.

If $\varepsilon_n^p n \to \infty$, then the liminf inequality follows from (17) of Lemma 4.8, while, the limsup inequality follows directly from

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

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 [8, Theorem 1.21] or Theorem 3.2.

Proof of Theorem 2.1. Note that, by [51, Theorem 13.2], with probability one the graph is eventually connected. For the remainder of the proof we assume n is chosen large enough so that the graph is connected. Let f_n be a minimizer of $\mathcal{E}_{n,\mathrm{con}}^{(p)}$ and $M \geq \max_{i=1,\dots,N} |y_i|$. If $\|f_n\|_{L^\infty(\mu_n)} > M$ then $\hat{f}_n$ defined by $\hat{f}_n = \max\{\min\{f_n, M\}, -M\}$ satisfies $\mathcal{E}_{n,\mathrm{con}}^{(p)}(\hat{f}_n) < \mathcal{E}_{n,\mathrm{con}}^{(p)}(f_n)$ hence f_n is not a minimizer. Thus, for n sufficiently large, $\|f_n\|_{L^\infty(\mu_n)} \leq M$, and we can restrict the minimization to the set of (f_n, μ_n) such that $\|f_n\|_{L^\infty(\mu_n)} \leq M$. This allows us to consider the setting of Theorem 4.7.

By the 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 $\mathcal{E}_{\infty,\mathrm{con}}^{(p)}$. Since the minimizer of $\mathcal{E}_{\infty,\mathrm{con}}^{(p)}$ is unique the convergence holds along the whole sequence.

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

Note that, in the above proof, if the graph is disconnected then any minimizer f_n of $\mathcal{E}_n^{(p)}$ can be redefined on any connected component that does not contain labeled data without changing the value of $\mathcal{E}_{n,\mathrm{con}}^{(p)}$. In particular, say $\{x_i\}_{i\in Z_n}$ where $Z_n \subset \{1, \dots, n\}$ is a connected component, and assume there is no labeled data in Z_n, i.e. $\min\{i \in Z_n\} > N$. Then, for a minimizer f_n of $\mathcal{E}_{n,\mathrm{con}}^{(p)}$ one can define $\hat{f}_n(x_i) = Q$ for $i \in Z_n$ and $\hat{f}_n(x_i) = f_n(x_i)$ otherwise. It follows that $\mathcal{E}_{n,\mathrm{con}}^{(p)}(\hat{f}_n) \leq \mathcal{E}_{n,\mathrm{con}}^{(p)}(f_n)$ and hence $\hat{f}_n$ is also a minimizer. Taking $Q \to \infty$ as $n \to \infty$ leads to a lack of compactness.

Note also that the connectivity of the graph is necessary in multiple places, for example Proposition 4.4. Indeed, Assumption **(A5)** implies that $\varepsilon_n \gg \|T_n - \mathrm{Id}\|_{L^\infty}$, since the connectivity radius of the graph can be bounded from above by $2\|T_n - \mathrm{Id}\|_{L^\infty}$ then, with probability one, we eventually have that ε_n is greater than the connectivity radius.

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

5. Improved Model. Let us restrict ourselves to when $p > d$. We recall that our main results thus far have been to show that constrained minimizers of $\mathcal{E}_n^{(p)}$, and the related penalized formulation $\mathcal{S}_n^{(p)}$, given by (9), are consistent as $n \to \infty$ only if the lower bound **(A5)** holds and $\frac{1}{n} \gg \varepsilon_n^p$. As remarked earlier (in Section 2), and demonstrated by our numerical experiments, this results in a narrow band of admissible ε_n for which the discrete models, $\mathcal{E}_{n,\mathrm{con}}^{(p)}$ and $\mathcal{S}_n^{(p)}$, are good approximations of the continuum models, $\mathcal{E}_{\infty,\mathrm{con}}^{(p)}$ and $\mathcal{S}_\infty^{(p)}$. In order to remove the upper bound, and hence widen the range of admissible ε_n, we propose a new model that is consistent for any $\varepsilon_n \to 0$ satisfying the lower bound **(A5)**.

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

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

Let $L = \min\{|x_i - x_j| \ : \ i \neq j, \text{ and } i, j \in \{1, \dots, N\}\}/2$ and $R_n = \min\{(1+\alpha)\varepsilon_n, L\}$ for $\alpha > 0$. The new

functional is defined by

$$\mathcal{F}^{(p)}_{n,\mathrm{con}}(f) = \begin{cases} \frac{1}{\varepsilon_n^p}\frac{1}{n^2}\sum_{i,j=1}^n W_{ij}|f(x_i)-f(x_j)|^p & \text{if } f \in C_n^{(R_n)} \\ \infty & \text{else.} \end{cases} \tag{24}$$

Since, for n sufficiently large, $R_n = (1+\alpha)\varepsilon_n$ in the sequel we will just define $R_n = (1+\alpha)\varepsilon_n$. We note that $\mathcal{F}^{(p)}_{n,\mathrm{con}}(f) \geq \mathcal{E}^{(p)}_{n,\mathrm{con}}(f)$ and for $f \in C_n^{(R_n)}$, $\mathcal{F}^{(p)}_{n,\mathrm{con}}(f) = \mathcal{E}^{(p)}_{n,\mathrm{con}}(f)$.

For asymptotic consistency we still need to require $p > d$, since only then is the limiting model $\mathcal{E}^{(p)}_{\infty,\mathrm{con}}$ 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 Assumptions **(A1) - (A8)**. Let f_n be a sequence of minimizers of $\mathcal{F}^{(p)}_{n,\mathrm{con}}$ defined in (24) with $R_n = (1+\alpha)\varepsilon_n$, $\alpha > 0$, and graph weights W_{ij} given by (5). 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 $\mathcal{E}^{(p)}_{\infty,\mathrm{con}}$ defined in (11) (with $\mathcal{E}^{(p)}_{\infty}$ defined in (10)).*

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, $\mathcal{F}^{(p)}_{n,\mathrm{con}}$ Γ-converges as $n \to \infty$ in the TL^p metric on the set $\{(\nu, g) : \nu \in \mathcal{P}(\Omega), \|g\|_{L^\infty(\nu)} \leq M\}$ to the functional $\mathcal{E}^{(p)}_{\infty,\mathrm{con}}$.*

We note that inequalities (17) of Lemma 4.8, and Proposition 4.4 hold for $\mathcal{F}^{(p)}_{n,\mathrm{con}}$ since $\mathcal{E}^{(p)}_{n,\mathrm{con}} \leq \mathcal{F}^{(p)}_{n,\mathrm{con}}$. We now turn to proving the liminf property and the existence of recovery sequence needed to show that $\mathcal{F}^{(p)}_{n,\mathrm{con}}$ Γ-converges in the TL^p topology to $\mathcal{E}^{(p)}_{\infty,\mathrm{con}}$.

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*

$$\mathcal{E}^{(p)}_{\infty,\mathrm{con}}(f) \leq \liminf_{n\to\infty} \mathcal{F}^{(p)}_{n,\mathrm{con}}(f_n). \tag{25}$$

Proof. Let (μ_n, f_n) be a convergent sequence in TL^p, such that f_n are uniformly bounded in $L^\infty(\mu_n)$, and $\liminf_{n\to\infty} \mathcal{F}^{(p)}_{n,\mathrm{con}}(f_n) < \infty$. 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 $B(x_i, (1+\alpha)\varepsilon_n)$ and $\|\mathrm{Id} - T_n\|_{L^\infty(\Omega)} < \alpha\varepsilon_n$ we have that $|T_n(x) - x_i| < (1+\alpha)\varepsilon_n$ for $x \in B(x_i, \varepsilon_n)$, hence $\tilde{f}_n := f_n \circ T_n = y_i$ on $B(x_i, \varepsilon_n)$ and consequently for $g_n := J_{\tilde{\varepsilon}_n} * \tilde{f}_n$ it holds that $g_n(x_i) = y_i$. Furthermore, note that $\|g_n\|_{L^\infty(\Omega)} \leq M$. By the bounds of Lemma 4.2 and Lemma 4.3, g_n is uniformly bounded in $W^{1,p}(\Omega')$ for any $\Omega' \subset\subset \Omega$ (Ω' compactly supported in Ω). 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 $i = 1, \dots, N$. Combining this with statement (17) of Lemma 4.8 yields (25). ∎

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*

$$\mathcal{E}^{(p)}_{\infty,\mathrm{con}}(f) \geq \limsup_{n\to\infty} \mathcal{F}^{(p)}_{n,\mathrm{con}}(f_n). \tag{26}$$

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

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 $i = 1, \dots, N$ then the proof of (26) is the same as the proof of (19). 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 $\delta_n \geq R_n$ converging to zero, which satisfies

$$\lim_{n\to\infty} \mathcal{E}^{(p)}_{\infty,\mathrm{con}}(g_n) = \mathcal{E}^{(p)}_{\infty,\mathrm{con}}(f). \tag{27}$$

We construct the sequence in the following way. Let θ be a cut-off function supported in $B(0, 1+\alpha)$. 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 $B(0, 1+\alpha)$, and $|\nabla\theta| < C$. 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,

$$\begin{aligned}\left|\mathcal{E}^{(p)}_{\infty,\mathrm{con}}(g_n) - \mathcal{E}^{(p)}_{\infty,\mathrm{con}}(f)\right| &\leq \sigma_\eta \int_\Omega \left||\nabla g_n(x)|^p - |\nabla f(x)|^p\right| \rho^2(x)\,\mathrm{d}x \\ &\leq \sigma_\eta \int_{B(x_1,(1+\alpha)\delta_n)} \left(|\nabla g_n(x)|^p + |\nabla f(x)|^p\right) \rho^2(x)\,\mathrm{d}x.\end{aligned}$$

We estimate

$$\int_{B(x_1,(1+\alpha)\delta_n)} |\nabla g_n|^p \rho^2\,\mathrm{d}x \leq 2^p \int_{B(x_1,(1+\alpha)\delta_n)} \left(|(f(x_1) - f(x))\nabla\theta_{\delta_n}(x - x_1)|^p + |\nabla f(x)|^p\right) \rho^2(x)\,\mathrm{d}x.$$

Using that $f \in C^{0,1-d/p}$ and furthermore, by the remark following Theorem 4 in Section 5.6.2 of [23] (where Theorem 4 is Morrey's inequality) we obtain

$$\begin{aligned}\int_{B(x_1,(1+\alpha)\delta_n)} |(f(x) - f(x_1))\nabla\theta_{\delta_n}(x - x_1)|^p \rho^2(x)\,\mathrm{d}x &\leq C_1 \delta_n^{p-d} \|\nabla f\|^p_{L^p(B(x_1,(1+\alpha)\delta_n))} \|\nabla\theta_{\delta_n}\|^p_{L^p(\mathbb{R}^d)} \\ &\leq C_1 \|\nabla f\|^p_{L^p(B(x_1,(1+\alpha)\delta_n))} \|\nabla\theta\|^p_{L^p(\mathbb{R}^d)}.\end{aligned}$$

Since $\lim_{n\to\infty} \int_{B(x_1,(1+\alpha)\delta_n)} |\nabla f(x)|^p\,\mathrm{d}x = 0$, by combining the inequalities above we conclude that (27) holds.

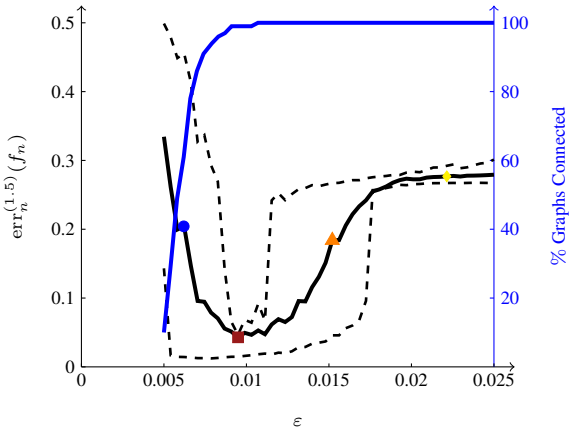
Generalizing to $N > 1$ is straightforward. ∎

6. Numerical Experiments. The results of Theorem 2.1 show that when $\varepsilon_n^p n \to 0$ then the solutions to the SSL problem (7) converge to a solution of the continuum constrained problem (11), while when $\varepsilon_n^p n \to \infty$ they degenerate as $n \to \infty$. However, in practice, for finite n, this does not provide precise guidance on what ε are appropriate. We investigate, via numerical experiments in 1D and 2D, the affect of ε on solutions to (7) in elementary examples. We also numerically compare the results with our improved model (24).

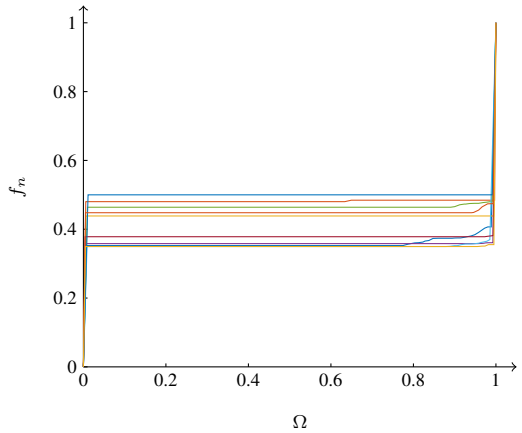
6.1. 1D Numerical Experiments. Let μ be the uniform measure on $[0,1]$ and consider η defined by $\eta(t) = 1$ if $t \leq 1$ and $\eta(t) = 0$ otherwise. We consider two different values of p: $p = 1.5$ and $p = 2$. The training set is $\{(0,0),(1,1)\}$, that is we condition on functions f_n taking the 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 $\mathcal{E}^{(1)}_{\infty,\mathrm{con}}$. For $p > 1$ the solution to the constrained limiting problem is $f^\dagger(x) = x$ (note that this is independent of p). Since $f^\dagger$ is continuous we can consider the following simple-to-compute notion of error:

$$\mathrm{err}_n^{(p)}(f_n) = \|f_n - f^\dagger\|_{L^p(\mu_n)}. \tag{28}$$

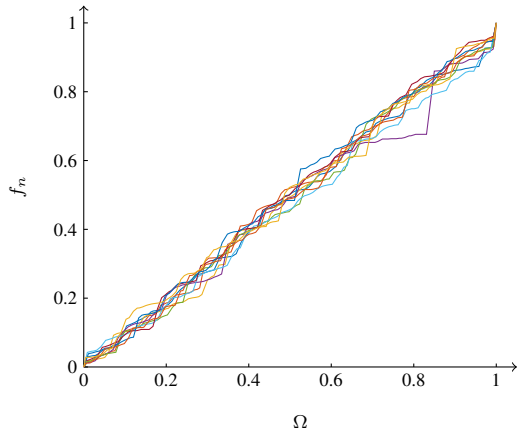
To find minimizers of (7) we use coordinate gradient descent. We enforce the constraints by choosing an initialisation that agrees with the training data and only updating coordinates that are not part of the training data set thereafter. The number of data points varies from $n = 80$ to $n = 5120$. For each n, ε and p we consider 100 different realizations of the random sample $\{x_i\}_{i=1}^n$ 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 discontinuities to arise and cause degeneracy. In Figure 2(a) and Figure 3(a) we plot the error as a function of ε for fixed $n = 1280$. We see clear



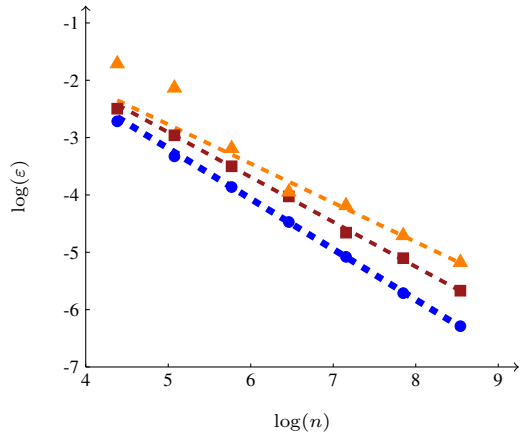
(a) Error (28) for $n = 1280$. The black line is the mean error, dashed lines are the 10% and 90% quantiles. The blue dot is the connectivity bound $\varepsilon_{\text{conn}}$, the red square is the optimal choice $\varepsilon_*^{(1.5)}$, and the orange triangle is the upper bound $\varepsilon_{\text{upper}}^{(1.5)}$. The blue line is the observed percentage of connected graphs for a given ε.



(b) We plot the functions output by the algorithm corresponding to nine realizations of the data for $n = 1280$ and $\varepsilon = 0.022$ (marked in yellow in Figure (a)).



(c) The output of the algorithm, f_n, for nine realizations of the data for $n = 1280$ and $\varepsilon = \varepsilon_*^{(1.5)}$.



(d) Orange triangles are $\varepsilon_{\text{upper}}^{(1.5)}$, red squares are $\varepsilon_*^{(1.5)}$, and blue dots are $\varepsilon_{\text{conn}}$. Lines show the linear fit over the last 5 points.

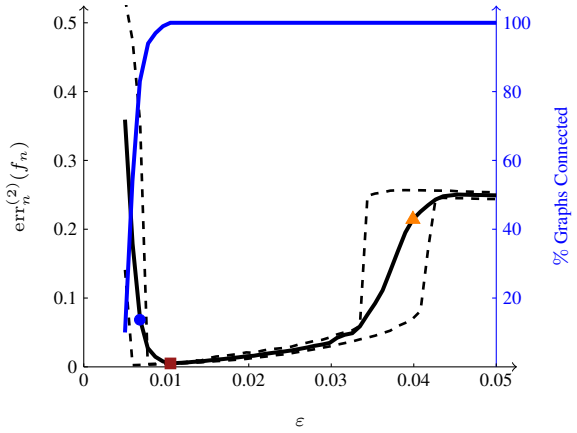
Fig. 2: 1D numerical experiments averaged over 100 realizations for (7) with $p = 1.5$.

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, see Figure 2(b), show that minimizers converge to a constant outside of the training data.

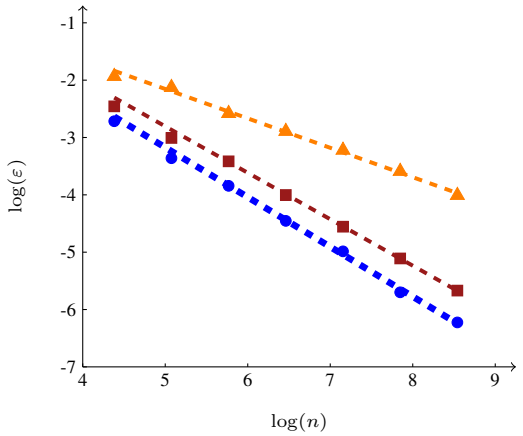
To measure how the transition points in ε, where minimizers change behavior, scale with n, we define the following:

(i) Given a realization $\{x_i^\omega\}_{i=1}^n$ let $\varepsilon_{\text{conn}}(n;\omega)$ be the connectivity radius for the particular realization, ω, that is $\varepsilon_{\text{conn}}(n;\omega)$ is the smallest ε such that the graph with weights $W_{ij} = \eta_\varepsilon(|x_i^\omega - x_j^\omega|)$ is connected. The value $\varepsilon_{\text{conn}}(n) = \frac{1}{M}\sum_{i=1}^M \varepsilon_{\text{conn}}(n;\omega_i)$ is the average connectivity radius. We considered $M = 100$ realizations.

(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 (7) with $\varepsilon_n = \varepsilon$; again averaged over $M = 100$ realizations.

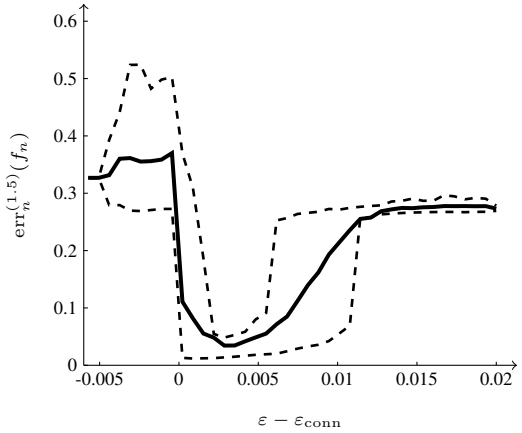


(a) Error (28) for $n = 1280$. The black line is the mean error, dashed lines are the 10% and 90% quantiles. The blue dot is the connectivity bound $\varepsilon_{\mathrm{conn}}$, the red square is the optimal choice $\varepsilon_*^{(2)}$ and the orange triangle is the upper bound $\varepsilon_{\mathrm{upper}}^{(2)}$. The blue line is the observed percentage of connected graphs for a given ε.

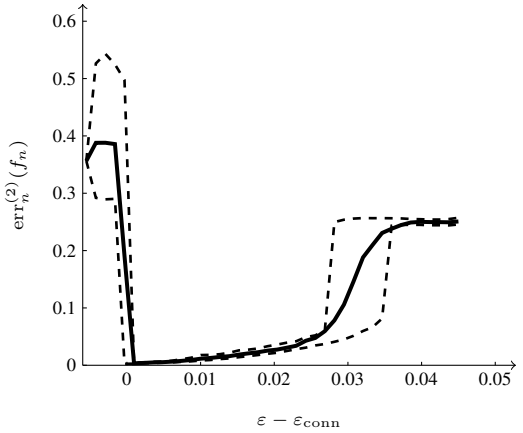


(b) Orange triangles are $\varepsilon_{\mathrm{upper}}^{(2)}$, red squares are $\varepsilon_*^{(2)}$, and blue dots are $\varepsilon_{\mathrm{conn}}$. Lines show the linear fit over the last 5 points.

Fig. 3: 1D numerical experiments averaged over 100 realizations for (7) with $p = 2$.



(a) Error (28) for $n = 1280$ and $p = 1.5$. The solid line is the mean error, the dashed lines are the 10% and 90% quantiles.

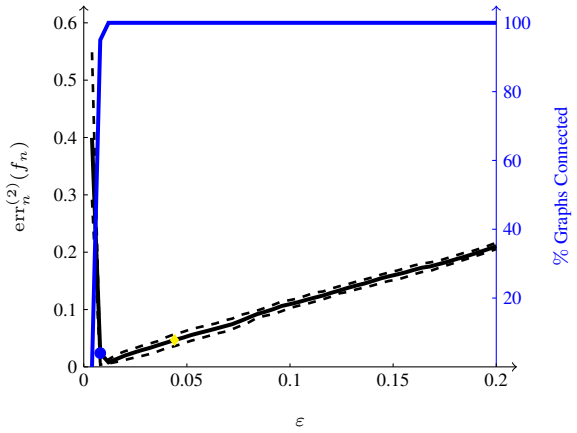


(b) Error (28) for $n = 1280$ and $p = 2$. The solid line is the mean error, the dashed lines are the 10% and 90% quantiles.

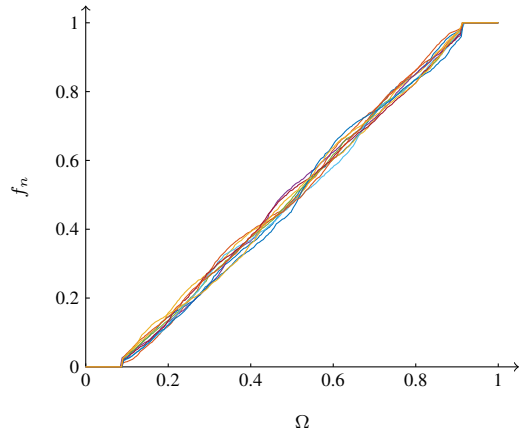
Fig. 4: Error shifted by connectivity radius using the same results as in Figures 2 and 3.

(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. As above the value is averaged over 100 realizations.

All of these points are highlighted in Figure 2(a) and Figure 3(a). In Figure 2(d) and Figure 3(b) 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 gives the



(a) Error (28) for $n = 1280$. The black line is the mean error, dashed lines are the 10% and 90% quantiles. The blue dot is the connectivity bound $\varepsilon_{\mathrm{conn}}$. The blue line is the observed percentage of connected graphs for a given ε.



(b) We plot the functions output from the algorithm corresponding to multiple realizations of the data for $n = 1280$ and $\varepsilon = 0.045$ (marked in yellow in Figure (a)).

Fig. 5: 1D numerical experiments averaged over 100 realizations for model (24) with $p = 2$ and $\alpha = 1$.

following scalings

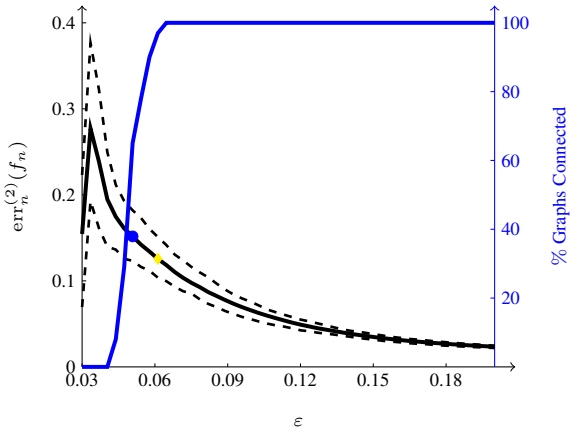
$$\varepsilon_*^{(1.5)}(n) \approx \frac{2.719}{n^{0.781}} \qquad \varepsilon_{\mathrm{upper}}^{(1.5)}(n) \approx \frac{1.905}{n^{0.683}}$$
$$\varepsilon_*^{(2)}(n) \approx \frac{3.472}{n^{0.810}} \qquad \varepsilon_{\mathrm{upper}}^{(2)}(n) \approx \frac{1.507}{n^{0.513}}$$
$$\varepsilon_{\mathrm{conn}}(n) \approx \frac{3.342}{n^{0.879}}.$$

We observe that asymptotic scaling established in Theorem 2.1 for $\varepsilon_{\mathrm{upper}}^{(p)}$ is $\frac{1}{n^{0.5}}$ for $p = 2$ and $\frac{1}{n^{0.667}}$ for $p = 1.5$, which is very close to our numerical results. The true scaling in the connectivity of the graph is $\frac{\ln(n)}{n}$, our numerical results behave approximately as $\frac{1}{n^{0.879}}$. We note that if we instead fit $\varepsilon_{\mathrm{conn}}(n) \approx c\left(\frac{\ln n}{n}\right)^{\alpha}$ we obtain $c = 1.266$ and $\alpha = 1.024$.

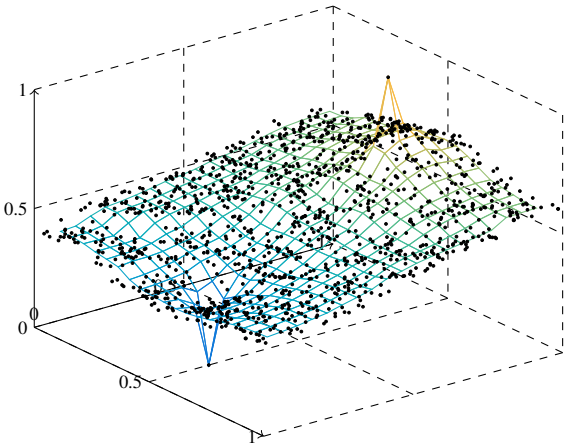
We observe that optimal choice $\varepsilon_*^{(p)}$ is quite close to the connectivity radius $\varepsilon_{\mathrm{conn}}(n)$. Choosing ε smaller than $\varepsilon_{\mathrm{conn}}(n)$ results in a large error due to trivial solutions, i.e. when the graph is disconnected minimizers are piecewise constant. To further investigate the proximity of the connectivity radius and the optimal choice of ε we plot in Figure 4 the error as the function of the size of ε relative to the connectivity radius. More precisely, we consider $\mathrm{err}_n^{(p)}(f_n; \varepsilon, \omega)$, where f_n is the minimizer of (7) for given ε and realization ω, as a function of $\varepsilon - \varepsilon_{\mathrm{conn}}(n; \omega)$ and then average over $M = 100$ realizations. We observe that, for both $p = 1.5$ and $p = 2$, the error is smallest when ε is quite close to the connectivity radius. The slight difference is that for $p = 1.5$ there is a short interval beyond the connectivity radius where the error is still decreasing.

Remark 6.1. Numerical experiments indicate the close proximity of the optimal epsilon to the connectivity radius, both for the original model and the improved model and both in 1D and 2D. This is not obvious since for ε small (i.e. close to the connectivity radius) the discrete p-Laplacian (i.e. the Euler–Lagrange equation of $\mathcal{E}_n^{(p)}(f)$) evaluated for a fixed smooth function f may fail to converge pointwise to the continuum p-Laplacian (the Euler–Lagrange equation of $\mathcal{E}_\infty^{(p)}(f)$) as $n \to \infty$. Namely as is shown in [55] for $p = 2$, if ε_n is small the variance can be large. Explaining the observed behavior of the error is an interesting open problem that we believe should be approached from the viewpoint of stochastic homogenization.

The improved model (24), for which we show results in Figure 5, is far more robust to the choice of ε. We plot the error as a function of ε for $n = 1280$ and we see a much larger range in the admissible choices of ε. Note that

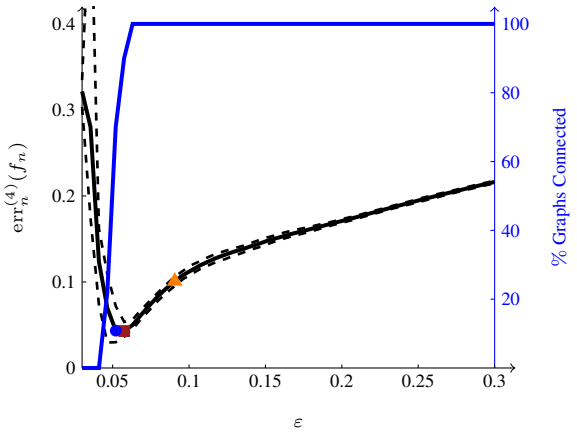


(a) Error (30) for $n = 1280$. The black line is the mean error, dashed lines are the 10% and 90% quantiles. The blue dot is the connectivity bound $\varepsilon_{\mathrm{conn}}$. The blue line is the observed percentage of connected graphs for given ε.

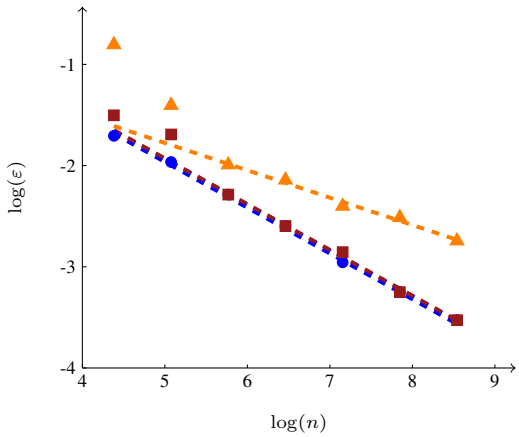


(b) We plot an example of a function output from the algorithm corresponding to $n = 1280$ and $\varepsilon = 0.06$ (marked in yellow in Figure (a)). The grid is to aid visualisation.

Fig. 6: 2D numerical experiments averaged over 100 realizations for (7) with $p = 2$.



(a) Error (29) for $n = 1280$. The black line is the mean error, dashed lines are the 10% and 90% quantiles. The blue dot is the connectivity bound $\varepsilon_{\mathrm{conn}}$, the red square is $\varepsilon_*^{(4)}$, and the orange triangle is the upper bound $\varepsilon_{\mathrm{upper}}^{(4)}$. The blue line is the observed percentage of connected graphs for a given ε.

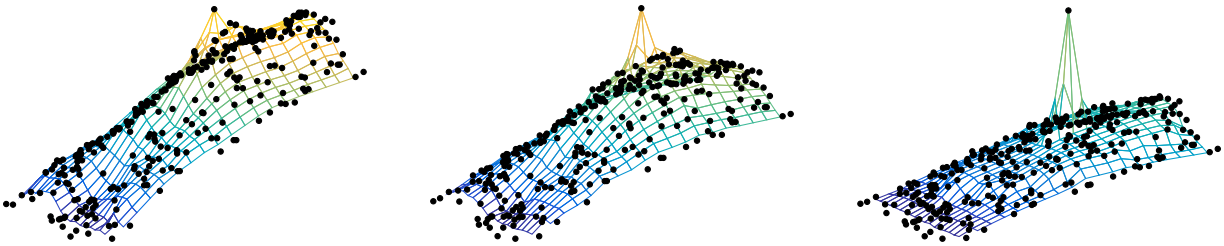


(b) Orange triangles are $\varepsilon_{\mathrm{upper}}^{(4)}$, red squares are $\varepsilon_*^{(4)}$, and blue dots are $\varepsilon_{\mathrm{conn}}$. Lines show the linear fit over the last 5 points.

Fig. 7: 2D numerical experiments averaged over 100 realizations for (7) with $p = 4$.

the horizontal axis covers a much larger range on Figure 5(a) compared to Figure 3(a). The comparison shows that model (7) does not produce a reasonable output when $\varepsilon \gtrsim 0.04$, while all outputs of (24) (when ε is larger than the connectivity radius) are close to the solution of the continuum model with constraints, i.e. $\mathcal{E}_{\infty,\mathrm{con}}^{(p)}$.

6.2. 2D Numerical Experiments. Let μ be the uniform measure on $\Omega = [0,1] \times [0,1]$, and $\eta(t) = 1$ if $|t| \leq 1$, $\eta(t) = 0$ otherwise. In 2D the critical value of p is $p = 2$, and we therefore choose to investigate $p = 2$ and $p = 4$. The training set is $x_1 = (0.2, 0.5)$, $x_2 = (0.8, 0.5)$, with labels $y_1 = 0$, $y_2 = 1$. In contrast to the 1D example the



(a) $\varepsilon = \varepsilon_*^{(4)}(n) \approx 0.0576$. (b) $\varepsilon = \varepsilon_{\text{upper}}^{(4)}(n) \approx 0.0906$. (c) $\varepsilon = 0.2$.

Fig. 8: Realizations of (7) with $p = 4$ and $n = 1280$ for a select choices of ε. Only the part of the domain near labeled point x_2 is shown. The grids are to aid visualization.

solution to the continuum problem (11) (in the well-posed regime) depends on p and furthermore cannot be solved analytically. To estimate the solution we discretized (11) on a uniform grid and ran a gradient descent algorithm to approximate the minimizer. In the case when $p = 4$ we plot our numerical approximation of the continuum minimizer to (11) in Figure 1(b). For $p > 2$ we define the error by

$$\text{err}_n^{(p)}(f_n) = \|f_n - f^{p,\dagger}\|_{L^p(\mu_n)} \tag{29}$$

where $f^{p,\dagger}$ minimizes (11). In the ill-posed case ($p \leq 2$) any constant function is a minimizer to the continuum problem, in which case we define the error as

$$\text{err}_n^{(p)}(f_n) = \inf_{c\in\mathbb{R}} \|f_n - c\|_{L^p(\mu_n)}. \tag{30}$$

To find minimizers of (7) for $p = 4$ we use coordinate gradient descent with constraints enforced similarly to when $d = 1$. For $p = 2$ we use the method of [70] that exactly solves the Euler Lagrange equation ($(L_n f)_i = 0$, where L_n is the graph Laplacian, for $i > 2$ and $f_1 = 0$, $f_2 = 1$). The number of data points varies from 80 to 5120. We use 100 different realizations of the data $\{x_i\}_{i=1}^n$ for each choice of n, ε and p.

Figure 6 shows the results in the ill-posed regime, for $p = 2$ and $n = 1280$. We observe that the solutions form spikes in order to satisfy the constraints. Spikes are present for all ε beyond the connectivity threshold, and grow as ε increases (recall that the solution to the continuum problem is a constant and therefore the error decreasing indicates convergence to a constant solution with spikes around the two training data points).

Figures 1 and 7 show the results for $p = 4$ which is in the well-posed range. Figure 1(a) presents the numerically computed discrete minimizer for the optimal radius $\varepsilon = \varepsilon_*^{(4)}$. We observe in Figure 7(a) that, similarly to 1D, for ε below the average connectivity radius, or when ε is large, the error is high, and is smallest for ε slightly larger than the connectivity threshold. In contrast to the 1D results we do notice that the transitions between the well-posed and ill-posed regime is gradual.

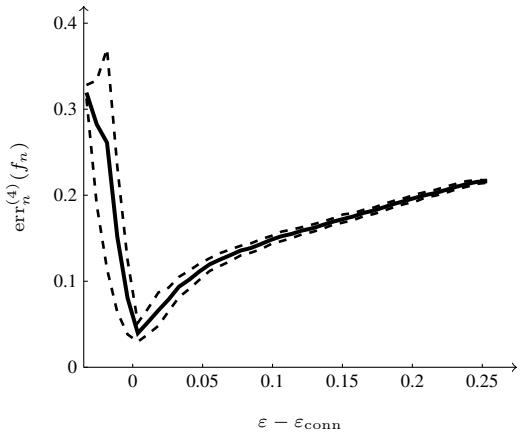
We find that the numerical scaling for $\varepsilon_*^{(p)}$, $\varepsilon_{\text{upper}}^{(p)}$, and $\varepsilon_{\text{conn}}$ (with the same definitions of quantities as in the 1D experiments in the previous subsection) are

$$\varepsilon_*^{(4)}(n) \approx \frac{1.394}{n^{0.452}} \qquad \varepsilon_{\text{upper}}^{(4)}(n) \approx \frac{0.654}{n^{0.270}} \qquad \varepsilon_{\text{conn}}(n) \approx \frac{1.368}{n^{0.452}}.$$

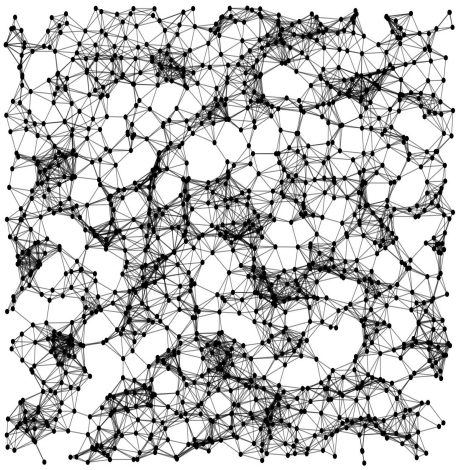
The connectivity radius should scale according to $\sqrt{\frac{\ln(n)}{n}}$, which is close to our observed rate of $n^{-0.452}$ (in fact when linear fitting $\ln\varepsilon$ to $\frac{\ln n}{n}$ one obtains $\varepsilon_{\text{conn}} \approx 0.829\left(\frac{\ln n}{n}\right)^{0.526}$). Our theoretical predictions give the scaling of the upper bound as $\varepsilon_{\text{upper}}^{(4)} \asymp n^{-0.25}$, close to our numerical rate of $n^{-0.270}$.

In Figure 8 we show instances of numerically computed minimizers of (7) for increasing values of ε. They show that the breakdown of the numerical approximation of the continuum solution happens via development of spikes.

As in the 1D examples we investigate the proximity of the optimal radius $\varepsilon_*^{(p)}(n;\omega)$ to the connectivity radius $\varepsilon_{\text{conn}}(n;\omega)$, where ω is the sample considered. In Figure 9(a) we plot the error, $\text{err}_n^{(p)}(f_n;\omega)$, against $\varepsilon - \varepsilon_{\text{conn}}(n;\omega)$

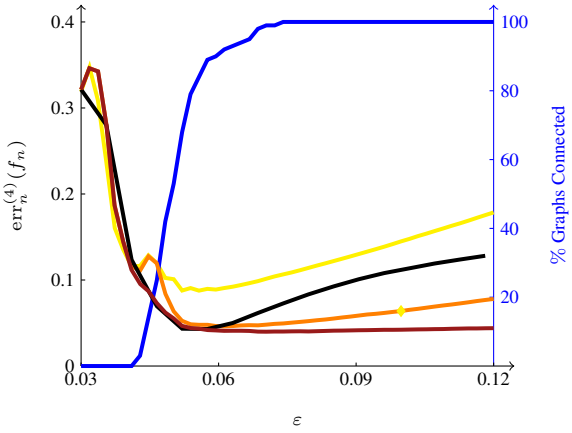


(a) Error (29) using the same results as in Figure 7. The solid line is the mean error, the dashed lines are the 10% and 90% quantiles.

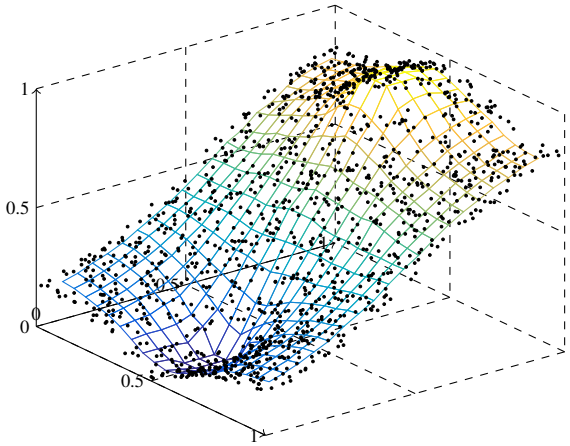


(b) Example graph in 2D for $\varepsilon = \varepsilon_*^{(4)}(n)$ and $n = 1280$.

Fig. 9: Error dependency on the connectivity radius and the graph for optimal ε, for $n = 1280$ and $p = 4$.



(a) The black line is the mean error (29) for model (7). The error for the improved model (24) with constraint radius $R_n = 2\varepsilon$ is in yellow, $R_n = \varepsilon$ is in orange, and $R_n = \varepsilon/2$ is in red.



(b) We plot an example of a function output from the algorithm corresponding to $n = 1280$ and $\varepsilon = 0.1$ for the constraint set of size ε (marked in yellow in Figure (a)). The grid is to aid visualisation.

Fig. 10: Experiments for improved model (24) with $n = 1280$ and $p = 4$ averaged over 100 realizations.

for $n = 1280$ and $p = 4$, averaged over 100 samples. The phenomenon we observe is similar to the 1D case; the error is large and highly variable for ε below the connectivity radius. There is a sharp transition to the well-posed regime, as soon as the graph is connected with the error increasing with ε. As we explain in Remark 6.1 it is an intriguing and important open problem to explain why the error is the smallest for rather coarse graphs (Figure 9(b)).

Our theoretical result in Section 5 showed that minimizers of the improved model, (24), converge as $n \to \infty$ to the correct solution if $1 \gg \varepsilon_n \gg (\ln n/n)^{1/d}$, regardless of how slowly $\varepsilon_n \to 0$. Here we numerically investigate two issues. One is how the error of the improved model depends on ε for fixed n. The other is to compare the observed error of the improved model to the original model. Recall that in the improved model we prove convergence when the labels are extended around the training set to balls of radius $(1 + \alpha)\varepsilon$ where $\alpha > 0$. This is needed in our proof to

ensure that spikes do not form. Here we numerically investigate if extending the labels to smaller balls, in particular choosing $\alpha \in (-1, 0]$, is sufficient to prevent spike formation. In Figure 10(a) we display the error for fixed $n = 1280$ and constraint ball radii 2ε, ε and $\varepsilon/2$. The numerics show that even radius $\varepsilon/2$ is sufficient to prevent spike formation and that it allows for better approximation of the continuum solution. We also observe that fixing the labels on larger sets can significantly negatively impact the accuracy of approximation. This issue is less pronounced for larger values of n, where the connectivity radius is small, and hence the constraint set can be chosen to be small. For example, for $n = 1280$ the connectivity radius is approximately $\varepsilon_{\text{conn}}(1280) \approx 0.05$, therefore a ball of radius $2\varepsilon_{\text{conn}}(1280)$ accounts for about 3% of the total domain. On the other hand the connectivity radius for $n = 5120$ is approximately $\varepsilon_{\text{conn}}(5120) \approx 0.03$ and therefore a ball of radius $2\varepsilon_{\text{conn}}(5120)$ accounts for about 1% of the total domain. Clearly, the larger n the smaller we can choose the constraint set and therefore the smaller the additional error.

Acknowledgements. The authors thank Matt Dunlop and Andrew Stuart for enlightening exchanges. The authors are grateful to Nicolás García Trillos for careful reading of the manuscript and insightful remarks.

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 (NIPS)*, pages 379–387, 2011.
[3] G. Alberti and G. Bellettini. A non-local anisotropic model for phase transitions: asymptotic behaviour of rescaled energies. *European Journal of Applied Mathematics*, 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)*, pages 953–960, 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 (NIPS)*, pages 129–136, 2007.
[7] A. Bertozzi, X. Luo, A. Stuart, and K. Zygalakis. Uncertainty quantification in graph-based classification of high dimensional data. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):568–595, 2018.
[8] A. Braides. Γ-*Convergence for Beginners*. Oxford University Press, 2002.
[9] 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.
[10] D. Burago, S. Ivanov, and Y. Kurylev. A graph discretization of the Laplace-Beltrami operator. *Journal of Spectral Theory*, 4(4):675–714, 2014.
[11] J. Calder. Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data. *arXiv preprint arXiv:1710.10364*, 2017.
[12] J. Calder. The game theoretic p-Laplacian and semi-supervised learning with few labels. *arXiv preprint arXiv:1711.10144*, 2017.
[13] J. Calder, S. Esedoḡlu, and A. O. Hero. A Hamilton-Jacobi equation for the continuum limit of nondominated sorting. *SIAM Journal on Mathematical Analysis*, 46(1):603–638, 2014.
[14] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.
[15] 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.
[16] G. Dal Maso. *An Introduction to* Γ-*Convergence*. Springer, 1993.
[17] Erik Davis and Sunder Sethuraman. Consistency of modularity clustering on random geometric graphs. *Ann. Appl. Probab.*, 28(4):2003–2062, 2018.
[18] M. Dunlop, D. Slepčev, A. M. Stuart, and M. Thorpe. Large data and zero noise limits of graph-based semi-supervised learning algorithms. *arXiv preprint arXiv:1805.09450*, 2018.
[19] 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.
[20] 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.
[21] 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.
[22] 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.
[23] L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 2010.
[24] C. Fefferman. Fitting a C^m-smooth function to data III. *Annals of Mathematics*, 2009.
[25] C. Fefferman, A. Israel, and G. K. Luli. Fitting a Sobolev function to data I. *Revista Matemática Iberoamericana*, 32(1):275–376, 2016.
[26] C. Fefferman, A. Israel, and G. K. Luli. Fitting a Sobolev function to data II. *Revista Matemática Iberoamericana*, 32(2):649–750, 2016.
[27] C. Fefferman, A. Israel, and G. K. Luli. Fitting a Sobolev function to data III. *Revista Matemática Iberoamericana*, 32(3):1039–1126, 2016.
[28] C. Fefferman and B. Klartag. Fitting a C^m-smooth function to data I. *Annals of Mathematics*, 169(1):315–346, 2009.
[29] C. Fefferman and B. Klartag. Fitting a C^m-smooth function to data II. *Revista Matemática Iberoamericana*, 25(1):49–273, 2009.
[30] I. Fonseca and G. Leoni. *Modern Methods in the Calculus of Variations: L^p Spaces*. Springer Science & Business Media, 2007.
[31] N. García-Trillos. Variational limits of k-NN graph based functionals on data clouds. *arXiv preprint arXiv:1607.00696*, 2016.
[32] N. García Trillos, M. Gerlach, M. Hein, and D. Slepčev. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs towards the Laplace-Beltrami operator. *arXiv preprint arXiv:1801.10108*, 2018.

[33] N. García Trillos, Z. Kaplan, T. Samakhoana, and D. Sanz-Alonso. On the consistency of graph-based Bayesian learning and the scalability of sampling algorithms. *arXiv preprint arXiv:1710.07702*, 2017.

[34] N. García Trillos and R. Murray. A new analytical approach to consistency and overfitting in regularized empirical risk minimization. *European J. Appl. Math.*, 28(6):886–921, 2017.

[35] N. García Trillos and D. Sanz-Alonso. Continuum limits of posteriors in graph Bayesian inverse problems. *SIAM J. Math. Anal.*, 50(4):4020–4040, 2018.

[36] 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.

[37] 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.

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

[39] 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.

[40] 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 Monograph Series*, pages 238–259. Institute of Mathematical Statistics, Beachwood, OH, 2006.

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

[42] 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.

[43] 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.

[44] G. Leoni. *A First Course in Sobolev Spaces*, volume 105. American Mathematical Society, 2009.

[45] 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.

[46] 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.

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

[48] 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.

[49] B. Osting and T. H. Reeb. Consistency of Dirichlet partitions. *SIAM J. Math. Anal.*, 49(5):4251–4274, 2017.

[50] B. Pelletier and P. Pudlo. Operator norm convergence of spectral clustering on level sets. *Journal of Machine Learning Research*, 12:385–416, 2011.

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

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

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

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

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

[56] 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.

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

[58] M. Thorpe, S. Park, S. Kolouri, G. K. Rohde, and D. Slepčev. A transportation L^p distance for signal analysis. *Journal of Mathematical Imaging and Vision*, 59(2):187–210, 2017.

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

[60] 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.

[61] 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.

[62] 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.

[63] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 2000.

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

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

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

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

[68] 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.

[69] 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.

[70] 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.