

The game theoretic p -Laplacian and semi-supervised learning with few labels

Jeff Calder*

July 22, 2019

Abstract

We study the game theoretic p -Laplacian for semi-supervised learning on graphs, and show that it is well-posed in the limit of finite labeled data and infinite unlabeled data. In particular, we show that the continuum limit of graph-based semi-supervised learning with the game theoretic p -Laplacian is a weighted version of the continuous p -Laplace equation. We also prove that solutions to the graph p -Laplace equation are approximately Hölder continuous with high probability. Our proof uses the viscosity solution machinery and the maximum principle on a graph.

1 Introduction

Large scale regression and classification problems are prominent today in data science and machine learning. In a typical problem, we are given a large collection of data, some of which is labeled, and our task is to extend the labels to the larger data set in some meaningful way. *Fully supervised* learning algorithms make use of only the labeled data. They may, for example, learn a parametrized function f on the ambient space that maximally agrees with the labeled data. Depending on the application, one might use support vector machines, or neural networks, for example. Fully supervised algorithms are generally only successful when there is a sufficiently diverse collection of labeled data to learn from. In many applications, labeled data requires human annotation, often by experts, and can be difficult and expensive to obtain.

On the other hand, due to the generally increasing availability of data, unlabeled data is essentially free. As such, there has recently been significant interest in *semi-supervised* learning algorithms that make use of both the labeled data, as well as the geometric or topological properties of the unlabeled data to achieve superior performance. Semi-supervised learning is used in many applications where a large amount of data is available, but only a very small subset is labeled. Examples include classification of medical images, natural

*Department of Mathematics, University of Minnesota. (jcalder@umn.edu)

language parsing, text recognition, website classification, protein sequencing to structure problems, and many others [1].

We consider semi-supervised learning on graphs. Here, we are given a weighted graph $G = (\mathcal{X}, \mathcal{W})$, where \mathcal{X} are the vertices, and $\mathcal{W} = \{w_{xy}\}_{x,y \in \mathcal{X}}$ are nonnegative edge weights. The weights are typically chosen so that $w_{xy} \approx 1$ when x and y are close, and $w_{xy} \approx 0$ when x and y are far apart (see Eq. (13)). The labeled data form a subset $\mathcal{O} \subset \mathcal{X}$ of the vertices called the *observation set*. Given an unknown real-valued function $g : \mathcal{X} \rightarrow \mathbb{R}$, and observations $g(x)$ for $x \in \mathcal{O}$, the goal of semi-supervised learning is to make predictions about the function g at the remaining vertices $x \in \mathcal{X} \setminus \mathcal{O}$.

Since there are infinitely many ways to extend the labels, the problem is ill-posed without some further assumptions. It is now standard to make the *semi-supervised smoothness assumption*, which says that we should extend the labeled data in the “smoothest” way possible, and the degree of smoothness should be locally proportional to the density of the graph. A widely used approach to the smoothness assumption is Laplacian regularization, which leads to the minimization problem

$$(1) \quad \min_{u: \mathcal{X} \rightarrow \mathbb{R}} \sum_{x,y \in \mathcal{X}} w_{xy}^2 (u(x) - u(y))^2 \quad \text{subject to } u(x) = g(x) \text{ for all } x \in \mathcal{O}.$$

Laplacian regularization was first proposed for learning tasks in [2], and has been extensively used since (see [3–6] and the references therein). It has also been applied to the related problem of manifold ranking [7–12], which has applications in object retrieval, for example.

Although Laplacian regularization is widely used and is successful for certain problems, it has been observed that when the number of unlabeled data points far exceeds the number of labeled points, the solution of (1) degenerates into a nearly constant label function u , with sharp spikes near each labeled data point $x \in \mathcal{O}$ [13, 14]. See Figure 1(a) for a simple example of this degeneracy. Thus, in the continuum limit as the number of unlabeled points tends to infinity while the number of labels is finite, the solutions of (1) do not continuously attain their boundary data on the set \mathcal{O} . In this sense, we say that Laplacian regularization is *ill-posed* in this regime. In general, we will say a learning algorithm is *well-posed* in the limit of infinite unlabeled data and finite labeled data if in the continuum limit the learned functions converge to a continuous function that attains the labels $u = g$ on \mathcal{O} continuously. If the continuum limit does not attain the labels continuously, then we say the algorithm is *ill-posed*.

To address the ill-posedness of Laplacian regularization, a class of ℓ_p -based Laplacian regularization models has recently been proposed [14]. For $1 \leq p \leq \infty$, ℓ_p -based Laplacian regularization leads to the optimization problem

$$(2) \quad \min_{u: \mathcal{X} \rightarrow \mathbb{R}} J_p(u) \quad \text{subject to } u(x) = g(x) \text{ for all } x \in \mathcal{O},$$

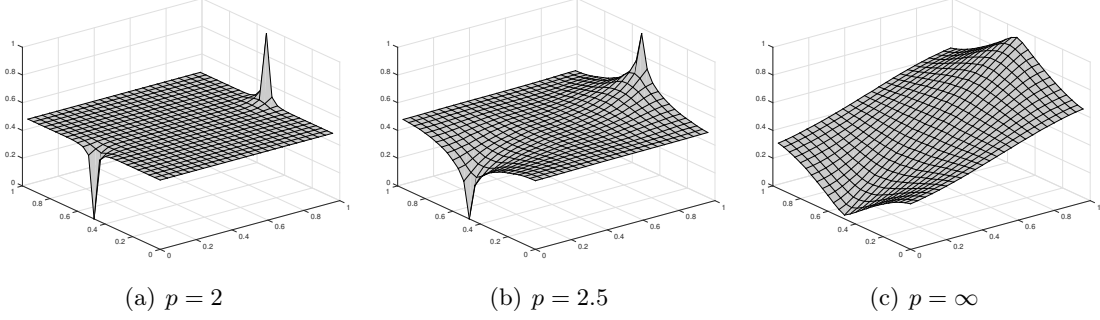


Figure 1: Example of ℓ_p -Laplacian regularized learning with (a) $p = 2$, (b) $p = 2.5$ and (c) $p = \infty$. In each case, the observed (labeled) data is $u(0, 0.5) = 0$ and $u(1, 0.5) = 1$, and the unlabeled data is a collection of $n = 10^5$ independent and uniformly distributed random variables on $[0, 1]^2$. The weights were set to $w_{xy} = 1$ if $|x - y| \leq 0.01$ and zero otherwise.

where

$$(3) \quad J_p(u) := \sum_{x, y \in \mathcal{X}} w_{xy} |u(x) - u(y)|^p \quad \text{for } 1 \leq p < \infty,$$

and

$$(4) \quad J_\infty(u) := \max_{x, y \in \mathcal{X}} \{w_{xy} |u(x) - u(y)|\}.$$

See Figure 1 for an illustration of ℓ_p -based Laplacian regularization. As the value of p increases, the smoothness (at least visually) of the minimizer u of (2) increases. The ℓ_p models are discussed in [14–16], and the $p = \infty$ case—called Lipschitz learning—was studied in [17, 18]. For $p = \infty$, the minimizer is not unique, even when the graph is connected. To see why, suppose we have a minimizer u of $J_\infty(u)$ that satisfies $w_{xy} |u(x) - u(y)| < J_\infty(u)$ for some fixed $x \in \mathcal{X}$ and all $y \in \mathcal{X}$. Then we can modify the value of $u(x)$ by a small amount $\delta > 0$ while ensuring that $J_\infty(u)$ is unchanged. This shows that in general there are an infinite number of minimizers of J_∞ . To resolve this issue, Kyng et al. [17] suggest to consider the minimizer whose gradient $Du := (w_{xy} |u(x) - u(y)|)_{x, y \in \mathcal{X}}$ is smallest in the lexicographic ordering on $\mathbb{R}^{|\mathcal{X}|^2}$. This amounts to minimizing the largest value of $Du(x, y)$, and then minimizing the second largest, and then the third largest, and so on. In the continuum setting, this is related to absolutely minimal Lipschitz extensions, which have a long history in analysis [19].

We note that the ℓ_p -based Laplacian regularization problem (2) is related in the continuum to the variational problem

$$(5) \quad \min_u \int_\Omega |\nabla u|^p dx.$$

Minimizers of (5) (for $1 \leq p < \infty$) satisfy the Euler-Lagrange equation

$$(6) \quad \Delta_p u := \operatorname{div}(|\nabla u|^{p-2} \nabla u) = 0.$$

The operator Δ_p is called the p -Laplacian, and solutions of (6) are called p -harmonic functions [20]. For $p > 2$, the p -Laplace equation (6) is degenerate, while for $1 \leq p < 2$ the operator is singular. Solutions of (6) can be interpreted in the weak (i.e., distributional) sense, or in the viscosity sense, and solutions are at most $C^{1,\alpha}$ when $p \neq 2$ [20]. Notice we can formally expand the divergence in (6) to obtain

$$\Delta_p u = |\nabla u|^{p-2} (\Delta u + (p-2) \Delta_\infty u),$$

where the ∞ -Laplacian Δ_∞ is given (when $\nabla u \neq 0$) by

$$(7) \quad \Delta_\infty u := \frac{1}{|\nabla u|^2} \sum_{i,j=1}^d u_{x_i x_j} u_{x_i} u_{x_j}.$$

Thus, any p -harmonic function satisfies the equation

$$(8) \quad \Delta u + (p-2) \Delta_\infty u = 0.$$

This equation is often called the *game-theoretic* or *homogeneous* p -Laplacian¹, since it arises in two player stochastic tug-of-war games [21, 22]. At the continuum level, the p -Laplace equation (6) is equivalent (at least formally) to the game theoretic version (8), while at the graph level, these two formulations of the *graph* p -Laplacian are different.

It was recently proved in [23] that in the limit of infinite unlabeled data and finite labeled data, solutions of the Lipschitz learning problem (i.e., (2) with $p = \infty$) converge to an ∞ -harmonic function that continuously attains the boundary condition $u = g$ on \mathcal{O} . In this sense, we can say that Lipschitz learning is *well-posed* in the limit of finite labeled data and infinite unlabeled data. However, the limiting ∞ -harmonic function is independent of the distribution of the unlabeled data (i.e., the algorithm “ignores” the unlabeled data), and hence Lipschitz learning is fully supervised in this limit [14, 23].

Along a similar thread, it was shown in [24] that ℓ_p -based Laplacian regularization is well-posed in the limit of finite labeled and infinite unlabeled data when $p > d$ and $\lim_{n \rightarrow \infty} n h_n^p = 0$, where d is dimension, h_n is the length scale used to define the weights in the graph (see (13)) and n is the number of vertices. This places a restriction on the length scale $h_n \ll (1/n)^{1/p}$ that may be undesirable in practice. We can see why this restriction is necessary with a basic energy scaling argument. If the n vertices in the graph are randomly sampled from \mathbb{R}^d and vertices are connected on a length scale of order $h_n > 0$, then each vertex is connected to on average $O(n h_n^d)$ neighbors, hence the sum defining $J_p(u)$ in (3)

¹Due to the identity $\Delta_p u = |\nabla u|^{p-2} (\Delta_1 u + (p-1) \Delta_\infty u)$, it is also common to call the quantity $\Delta_1 u + (p-1) \Delta_\infty u$ the game theoretic p -Laplacian [21]

has on the order of $n^2 h_n^d$ terms. If the function u is a non-constant smooth labeling of the data (i.e., the well-posed regime), then each term in the sum contributes $O(h_n^p)$ to the total, hence the energy scales like $J_p(u) \sim n^2 h_n^{d+p}$. On the other hand, if u is constant, except at L labels, then the only nonzero terms in the sum correspond to connections to labeled points, and the size of each term is $O(1)$. Hence, the energy of a constant labeling scales like $J_p(u) \sim L n h_n^d$. Of course, we want the energy of a constant labeling to be much larger than a non-constant (i.e., $L n h_n^d \gg n^2 h_n^{d+p}$) to ensure the algorithm is well-posed. This reduces to the requirement that $L \gg n h_n^p$. In the case that L is constant in n , then we need $n h_n^p \rightarrow 0$ as $n \rightarrow \infty$. Let us mention that the authors of [24] showed how to fix this issue by modifying the model (2) so that the boundary condition $u = g$ is required to hold on the dilated set $\mathcal{O} + B(0, 2h_n)$. In this case, the length scale restriction $n h_n^p \rightarrow 0$ is not required, since we are labeling $L \sim n h_n^d \gg n h_n^p$ data points (recall $p > d$). The disadvantage of the improved model is that it now requires an infinite number of labels, since $n h_n^d \rightarrow \infty$ is required for almost sure connectivity of the graph.

In machine learning, it is often assumed that while data may be presented in a very high dimensional space \mathbb{R}^D , most types of data lie close to some submanifold of some much smaller dimension. This is called the *manifold assumption*, and the dimension $d \ll D$ of the manifold is often called the *intrinsic dimension* of the data. We note that in the requirement $p > d$ in the well-posedness result from [24], the value of d should be interpreted as this intrinsic dimension, as opposed to extrinsic dimension D . The intrinsic dimension d can be much larger than $d = 2$ in practice. For example, some of the digits in the MNIST dataset² are estimated to have intrinsic dimension between $d = 12$ and $d = 14$ [26, 27], and the MNIST dataset has very low complexity compared to modern problems, such as ImageNet.⁴ While the Laplacian regularization ($p = 2$) problem can be solved efficiently [29, 30], to our knowledge there are no algorithms in the literature for solving the ℓ_p -based Laplacian regularization problem (2) efficiently and to scale, especially when $p \gg 2$. El alaoui et al. [14] suggest to use Newton’s method, while Kyng et al. [17] suggest convex programming, which requires very high accuracy for large p , but neither has been implemented and tested for large scale problems. To see why it may be difficult to solve (2) when p is large, we recall that the convergence rate for optimization algorithms (such as gradient descent or Newton’s method) for solving $\min_u F(u)$ depends on the Lipschitz norm of the gradient

$$(9) \quad \alpha := \sup_{u \neq v} \frac{\|\nabla F(u) - \nabla F(v)\|}{\|u - v\|}.$$

When α is very large iterative methods are slow to converge [31], and we say the problem is poorly conditioned. This is often manifested as an ill-conditioning of the Hessian matrix $\nabla^2 F$. It appears that for ℓ_p -based learning (2), $\alpha = \alpha(p)$ grows exponentially with p , suggesting the problem is challenging to solve for larger values of p .

²The MNIST dataset [25] consists of 70,000 black and white 28x28 pixel images of handwritten digits.

⁴The ImageNet dataset [28] consists of over 14 million natural images belonging to over 20,000 categories.

In this paper we study the game theoretic *graph* p -Laplacian as a regularizer for semi-supervised learning on graphs with very few labeled vertices. We show that semi-supervised learning with the game theoretic p -Laplacian is well-posed in the limit of finite labeled data and infinite unlabeled data when $p > d$. In particular, we show in this limiting regime that the solutions of the game theoretic graph p -Laplacian converge to the unique viscosity solution of a weighted p -Laplace equation, and the boundary values $u = g$ on \mathcal{O} are attained continuously in this limit. As in (8), the game theoretic graph p -Laplacian is a linear combination of the graph 2-Laplacian and the graph ∞ -Laplacian, with the hyperparameter p appearing as a coefficient. Therefore, the game theoretic p -Laplacian is expected to be well conditioned numerically for all values of p , since the condition number α from (9) can be bounded independently of p (also, see Remark 4 below). Furthermore, the continuum well-posedness of the game-theoretic p -Laplacian does not require any upper bound on the length scale h_n (only $h_n \rightarrow 0$ is required), which may be more desirable in practice. This work suggests that regularization with the game theoretic p -Laplacian may be useful in semi-supervised learning with few labels. We describe our main results and outline the paper in Section 1.1 below.

1.1 Main results

We now describe our setup and main results. We take periodic boundary conditions and work on the flat Torus $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$. Let $X_n = \{x_1, \dots, x_n\}$ be a sequence of independent and identically distributed random variables on \mathbb{T}^d with probability density $f \in C^2(\mathbb{T}^d)$. We assume that $f > 0$ on \mathbb{T}^d . Let $\mathcal{O} \subset \mathbb{T}^d$ be a fixed finite collection of points and set

$$(10) \quad \mathcal{X}_n := X_n \cup \mathcal{O}.$$

The points \mathcal{X}_n will form the vertices of our graph. To select the edge weights, let $\Phi : [0, \infty) \rightarrow [0, \infty)$ be a C^2 function satisfying

$$(11) \quad \begin{cases} \Phi(s) \geq 1, & \text{if } s \in (0, 1) \\ \Phi(s) = 0, & \text{if } s \geq 2. \end{cases}$$

We also assume that there exists $s_0 \in (0, 2)$ and $\theta > 0$ such that

$$(12) \quad s\Phi(s) + \theta(s - s_0)^2 \leq s_0\Phi(s_0) \text{ for all } 0 \leq s \leq 2.$$

Select a length scale $h_n > 0$ and define

$$(13) \quad w_n(x, y) := \Phi\left(\frac{|x - y|}{h_n}\right), \text{ and } d_n(x) = \sum_{y \in \mathcal{X}_n} w_n(x, y).$$

Here, $|x - y|$ denotes the distance on the flat torus \mathbb{T}^d . Let $\mathcal{W}_n = \{w_n(x, y)\}_{x, y \in \mathcal{X}_n}$ and let $\mathcal{G}_n = (\mathcal{X}_n, \mathcal{W}_n)$ be the graph with vertices \mathcal{X}_n and edge weights \mathcal{W}_n . The graph 2-Laplacian

is given by

$$(14) \quad L_n^2 u(x) := \sum_{y \in \mathcal{X}_n} w_n(x, y)(u(y) - u(x)),$$

while the graph ∞ -Laplacian is given by

$$(15) \quad L_n^\infty u(x) := \max_{y \in \mathcal{X}_n} w_n(x, y)(u(y) - u(x)) + \min_{y \in \mathcal{X}_n} w_n(x, y)(u(y) - u(x)).$$

We define the game theoretic graph p -Laplacian [32] by

$$(16) \quad L_n^p u(x) := \frac{1}{d_n(x)} L_n^2 u(x) + \lambda(p-2) L_n^\infty u(x),$$

where

$$(17) \quad \lambda = \frac{C_2}{s_0^2 \Phi(s_0) C_1},$$

$C_1 = \int_{B(0,2)} \Phi(|z|) dz$, and $C_2 = \frac{1}{2} \int_{B(0,2)} \Phi(|z|) z_1^2 dz$. The choice of λ ensures that for smooth functions φ , $L_n^p \varphi = 0$ is consistent (as $n \rightarrow \infty$ and $h_n \rightarrow 0$) with the weighted p -Laplacian

$$(18) \quad \operatorname{div}(f^2 |\nabla \varphi|^{p-2} \nabla \varphi) = |\nabla \varphi|^{p-2} (\operatorname{div}(f^2 \nabla \varphi) + (p-2) f^2 \Delta_\infty \varphi).$$

Let $g : \mathcal{O} \rightarrow \mathbb{R}$ be given labels and let $u_n : \mathcal{X}_n \rightarrow \mathbb{R}$ be the solution of the game theoretic p -Laplacian boundary value problem

$$(19) \quad \left. \begin{array}{ll} L_n^p u_n &= 0 \quad \text{in } X_n \\ u_n &= g \quad \text{in } \mathcal{O}. \end{array} \right\}$$

Our main result is the following continuum limit.

Theorem 1. *Let $d < p < \infty$, and suppose that $h_n \rightarrow 0$ such that*

$$(20) \quad \lim_{n \rightarrow \infty} \frac{n h_n^q}{\log(n)} = \infty,$$

where $q = \max\{d+4, 3d/2\}$. Then with probability one

$$(21) \quad u_n \longrightarrow u \quad \text{uniformly as } n \rightarrow \infty,$$

where $u \in C^{0, \frac{p-d}{p-1}}(\mathbb{T}^d)$ is the unique viscosity solution of the weighted p -Laplace equation

$$(22) \quad \left. \begin{array}{ll} \operatorname{div}(f^2 |\nabla u|^{p-2} \nabla u) &= 0 \quad \text{in } \mathbb{T}^d \setminus \mathcal{O} \\ u &= g \quad \text{on } \mathcal{O}. \end{array} \right\}$$

Theorem 1 proves that graph-based semi-supervised learning with the game theoretic p -Laplacian is well-posed in the limit of finite labeled data and infinite unlabeled data when $p > d$. We prove in Section 5 that weak distributional solutions of (22) are equivalent to viscosity solutions, so we can also interpret u as the unique weak solution of (22). By regularity theory for weak solutions of the weighted p -Laplace equation [33, 34], we have $u \in C_{loc}^{1,\alpha}(\mathbb{T}^d \setminus \mathcal{O})$ and if f is smooth then $u \in C^\infty(B(x, r))$ near any point where $\nabla u(x) \neq 0$.

A key step in proving Theorem 1 is a discrete regularity result that is useful to state independently.

Theorem 2. *For every $0 < \alpha < (p - d)/(p - 1)$ there exists $C, \delta > 0$ such that*

$$(23) \quad \mathbb{P}[\forall x, y \in \mathcal{X}_n, |u_n(x) - u_n(y)| \leq C(|x - y|^\alpha + h_n^\alpha)] \geq 1 - \exp(-\delta n h_n^q + C \log(n)),$$

where $q = \max\{d + 4, 3d/2\}$.

Theorem 2 quantifies the way in which the game theoretic p -Laplacian regularizes the learning algorithm; it asks that u_n is approximately Hölder continuous.

Several remarks are in order.

Remark 1. *Notice that (22) is the Euler-Lagrange equation for the variational problem*

$$\min_{u: \mathbb{T}^d \rightarrow \mathbb{R}} \int_{\mathbb{T}^d} f^2 |\nabla u|^p dx$$

subject to $u = g$ on \mathcal{O} . This variational problem is also the continuum limit of ℓ_p -based Laplacian regularization [14]. Hence, ℓ_p -based Laplacian regularization and regularization via the game theoretic p -Laplacian, while different models at the graph level, are identical in the continuum.

Remark 2. *We expect that when $p \leq d$, u_n converges pointwise to a constant $c \in \mathbb{R}$. It would be interesting to prove this and determine the constant c .*

Remark 3. *The proof of Theorem 1 uses techniques from the theory of viscosity solutions of partial differential equations [35]. The same ideas can be used to prove a similar result for the graph p -Laplacian boundary value problem*

$$(24) \quad \left. \begin{aligned} \sum_{y \in \mathcal{X}_n} w_n(x, y) |u_n(x) - u_n(y)|^{p-2} (u_n(x) - u_n(y)) &= 0 & \text{if } x \in X_n \\ u_n(x) &= g(x) & \text{if } x \in \mathcal{O}, \end{aligned} \right\}$$

which corresponds to ℓ_p -based Laplacian regularization. The main difference in the proof is the consistency result, which we include for (24) in the appendix for completeness. This problem was studied very recently with variational (as opposed to PDE) techniques by Slepčev and Thorpe [24]. In this setting, there is an additional requirement that $\lim_{n \rightarrow \infty} n h_n^p = 0$, which places an upper bound on the length scale $h_n \ll 1/n^{1/p}$ that is not present in the game theoretic model.

Remark 4. *Theorem 1 suggests that the game theoretic p -Laplacian can be useful for regularization in graph-based semi-supervised learning with small amounts of labeled data. We note that the constituent terms in the game theoretic p -Laplacian—the graph 2-Laplacian and ∞ -Laplacian—are well-understood and efficient algorithms exist for solving both [17, 29]. Furthermore, on uniform grids (i.e., the PDE-version), there are efficient algorithms for the game theoretic p -Laplacian, such as the semi-implicit scheme of Oberman [36]. This suggests the game theoretic p -Laplacian on a graph can be solved efficiently and to scale, making it a good choice for semi-supervised learning with few labels.*

Remark 5. *If instead of working on the Torus \mathbb{T}^d , we take our unlabeled points to be sampled from a domain $X_n \subset \Omega \subset \mathbb{R}^d$, then we expect Theorem 1 to hold under similar hypotheses, with the additional boundary condition*

$$(25) \quad \frac{\partial u}{\partial \nu} = 0 \quad \text{on } \partial\Omega.$$

Let us outline briefly how we expect the proof to change. The Neumann condition (25) will make an appearance in the consistency result for the graph Laplacian (Theorem 5), which near the boundary $\partial\Omega$ will involve, for the 2-Laplacian, the integral

$$(26) \quad L_n^2 u(x) \approx n \int_{B(x, 2h_n) \cap \Omega} \Phi(h_n^{-1}|x-y|) (u(y) - u(x)) dy.$$

Taylor expanding u we obtain an expression of the form

$$L_n^2 u(x) \approx C n h_n^{d+1} \left(\frac{\text{dist}(x, \partial\Omega)}{h} - 1 \right) \frac{\partial u}{\partial \nu} + O(n h^{d+2}).$$

A similar statement holds for the ∞ -Laplacian. The remainder of the proof will be the same; the final difference is that we will require uniqueness of viscosity solutions to the weighted p -Laplace equation with Neumann boundary conditions. We refer the reader to the user's guide [35, Theorem 7.5] for a comparison principle for the generalized Neumann problem. Uniqueness of viscosity solutions will require some boundary regularity, and we expect $\partial\Omega \in C^1$ to be sufficient.

This paper is organized as follows. In Section 2 we recall the maximum principle on a graph and prove that (19) is well-posed. In Section 3 we prove consistency for the game theoretic p -Laplacian for smooth functions. In Section 4 we prove a discrete regularity result (Theorem 2) for the game theoretic p -Laplacian. Finally, in Section 5 we prove our main consistency result, Theorem 1.

2 The maximum principle

Our main tool is the maximum principle. We recall here the maximum principle on a graph, which was proved in a similar setting in [32]. Our setting is slightly more general, so we include a proof here for completeness.

We first introduce some notation. We say that y is adjacent to x whenever $w_n(x, y) > 0$. We say that the graph $\mathcal{G}_n = (\mathcal{X}_n, \mathcal{W}_n)$ is connected to $\mathcal{O} \subset \mathcal{X}_n$ if for every $x \in \mathcal{X}_n \setminus \mathcal{O}$ there exists $y \in \mathcal{O}$ and a path from x to y consisting of adjacent vertices.

Theorem 3 (Maximum principle). *Assume the graph \mathcal{G}_n is connected to \mathcal{O} . Let $p \geq 2$ and suppose $u, v : \mathcal{X}_n \rightarrow \mathbb{R}$ satisfy*

$$(27) \quad L_n^p u(x) \geq L_n^p v(x) \quad \text{for all } x \in X_n.$$

Then

$$(28) \quad \max_{\mathcal{X}_n}(u - v) = \max_{\mathcal{O}}(u - v).$$

In particular, if $u \leq v$ on \mathcal{O} , then $u \leq v$ on \mathcal{X}_n .

Proof. Let $x \in X_n$ be a point at which $u - v$ attains its maximum. If $x \in \mathcal{O}$, then we are done, so suppose that $x \in X_n$. Since $u - v$ attains its maximum at x we have

$$w_n(x, y)(u(y) - u(x)) \leq w_n(x, y)(v(y) - v(x)) \quad \text{for all } y \in \mathcal{X}_n.$$

If any of the inequalities above were strict, we would have $L_n^p u(x) < L_n^p v(x)$ which contradicts the assumption given in (27). Therefore

$$u(y) - v(y) = u(x) - v(x) \quad \text{whenever } w_n(x, y) > 0.$$

That is, $u - v$ also attains its maximum at every vertex that is adjacent to x . Since the graph is connected to \mathcal{O} , we can find a path from x to \mathcal{O} consisting of adjacent vertices. Therefore $u - v$ attains its maximum somewhere on \mathcal{O} , which completes the proof. \square

Remark 6. Notice the exact form of L_n^p is not used directly in the proof of Theorem 3. All that is required is that $L_n^p u(x)$ is a strictly increasing function of $w_n(x, y)(u(y) - u(x))$ for every $y \in \mathcal{X}_n$ with $y \neq x$. Finite difference schemes with this property are called monotone in the PDE literature [37].

We now establish existence and uniqueness of a solution to (19).

Theorem 4. *Assume \mathcal{G}_n is connected to \mathcal{O} and let $p \geq 2$. For any $g : \mathcal{O} \rightarrow \mathbb{R}$, there exists a unique solution $u_n : \mathcal{X}_n \rightarrow \mathbb{R}$ of (19). Furthermore, u satisfies the a priori estimate*

$$(29) \quad \min_{\mathcal{O}} g \leq u_n \leq \max_{\mathcal{O}} g.$$

Proof. We use the Perron method to prove existence. Define

$$\mathcal{F} := \left\{ v : \mathcal{X}_n \rightarrow \mathbb{R} : L_n^p v \geq 0 \text{ in } X_n \text{ and } \min_{\mathcal{O}} g \leq v \leq g \text{ in } \mathcal{O} \right\}$$

and for $x \in \mathcal{X}_n$ set

$$u_n(x) := \sup_{v \in \mathcal{F}} v(x).$$

Note that $v \equiv \min_{\mathcal{O}} g$ belongs to \mathcal{F} , so \mathcal{F} is nonempty. Furthermore, $w \equiv \max_{\mathcal{O}} g$ satisfies $L_n^p w \equiv 0$, and so by Theorem 3, $v \leq w$ for all $v \in \mathcal{F}$. It follows that u_n satisfies (29).

We now show that $L_n^p u_n \geq 0$ in X_n . Fix $x \in X_n$ and let $v_k \in \mathcal{F}$ be a sequence of functions such that $v_k(x) \rightarrow u_n(x)$ as $k \rightarrow \infty$. Since $\min_{\mathcal{O}} g \leq v_k \leq \max_{\mathcal{O}} g$ we can pass a subsequence, if necessary, so that $v_k \rightarrow v$ on \mathcal{X}_n for some $v : \mathcal{X}_n \rightarrow \mathbb{R}$. By continuity $L_n^p v \geq 0$ on X_n and $\min_{\mathcal{O}} g \leq v \leq g$ in \mathcal{O} . Therefore $v \in \mathcal{F}$, and so $u_n \geq v$ on \mathcal{X}_n and $u_n(x) = v(x)$. It follows that $L_n^p u_n(x) \geq L_n^p v(x) \geq 0$.

We now show that $L_n^p u_n \leq 0$ in X_n . Fix $x \in X_n$ and assume to the contrary that $L_n^p u_n(x) > 0$. Define

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

For $\varepsilon > 0$ define $v_\varepsilon := u_n + \varepsilon w$. By continuity, we can choose $\varepsilon > 0$ sufficiently small so that $L_n^p v_\varepsilon(x) \geq 0$. By definition we have $L_n^p v_\varepsilon(y) \geq L_n^p u_n(y) \geq 0$ for all $y \neq x$. Therefore $L_n^p v_\varepsilon \geq 0$ in X_n for sufficiently small $\varepsilon > 0$. Hence $v_\varepsilon \in \mathcal{F}$ and so $u_n(x) \geq v_\varepsilon(x) = u_n(x) + \varepsilon$, which is a contradiction.

Finally we show that $u_n = g$ on \mathcal{O} . By definition, we have that $u_n \leq g$ on \mathcal{O} . Assume to the contrary that $u_n(x) < g(x)$ for some $x \in \mathcal{O}$. Then as above we can set $v_\varepsilon = u_n + \varepsilon w$, and we find that $v_\varepsilon \in \mathcal{F}$ for $\varepsilon > 0$ sufficiently small, which is a contradiction. Uniqueness follows directly from Theorem 3. \square

3 Consistency

We now prove consistency for the game theoretic p -Laplacian applied to smooth test functions. This is split into two steps: In Section 3.2 we prove a consistency result for the graph 2-Laplacian and in Section 3.3 we prove a consistency result for the graph ∞ -Laplacian. We note there are numerous consistency results in the literature for the graph 2-Laplacian [38–44], but we require a slightly different notion of consistency to use the viscosity solution framework and prove discrete regularity.

3.1 Concentration of measure

Graph Laplacians are sums of *i.i.d.* random variables, and to prove consistency we need to control the fluctuations of the graph Laplacian about its mean. Our main tool for proving consistency is a standard concentration of measure result referred to as Bernstein's inequality [48], which we recall now for the reader's convenience. For Y_1, \dots, Y_n *i.i.d.* with variance $\sigma^2 = \mathbb{E}((Y_i - \mathbb{E}[Y_i])^2)$, if $|Y_i| \leq M$ almost surely for all i then Bernstein's inequality

states that for any $t > 0$

$$(30) \quad \mathbb{P} \left(\left| \sum_{i=1}^n Y_i - \mathbb{E}(Y_i) \right| > t \right) \leq 2 \exp \left(-\frac{t^2}{2n\sigma^2 + 4Mt/3} \right).$$

Since the graph Laplacian depends only on nearby nodes in the graph, the variance σ^2 is much smaller than the absolute bound M , which makes the Bernstein inequality far tighter (for small t) than other concentration inequalities, such as the Hoeffding inequality [45], which only depends on the size of M .

The following lemma converts the Bernstein inequality into a result that is directly useful for graph Laplacians.

Lemma 1. *Let $Y_1, Y_2, Y_3, \dots, Y_n$ be a sequence of i.i.d random variables on \mathbb{R}^d with Lebesgue density $f : \mathbb{R}^d \rightarrow \mathbb{R}$, let $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ be bounded and Borel measurable with compact support in a bounded open set $\Omega \subset \mathbb{R}^d$, and define*

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

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

$$(31) \quad \mathbb{P}(|Y - \mathbb{E}(Y)| > \|f\|_\infty \|\psi\|_\infty n |\Omega| \lambda) \leq 2 \exp \left(-\frac{1}{4} \|f\|_\infty n |\Omega| \lambda^2 \right),$$

where $\|\psi\|_\infty = \|\psi\|_{L^\infty(\Omega)}$, and $|\Omega|$ denotes the Lebesgue measure of Ω .

Proof. The proof is a direct application of the Bernstein inequality (30). We compute $M \leq \|\psi\|_\infty$ and

$$\sigma^2 \leq \mathbb{E}(\psi(Y_i)^2) = \int_{\Omega} \psi(y)^2 f(y) dy \leq |\Omega| \|\psi\|_\infty^2 \|f\|_\infty.$$

Therefore, Bernstein's inequality yields

$$(32) \quad \mathbb{P}(|Y - \mathbb{E}(Y)| > t) \leq 2 \exp \left(-\frac{t^2}{2n|\Omega| \|\psi\|_\infty^2 \|f\|_\infty + 4\|\psi\|_\infty t/3} \right),$$

for any $t > 0$. Setting $t = \|f\|_\infty \|\psi\|_\infty n |\Omega| \lambda$ for $\lambda > 0$ we have

$$(33) \quad \mathbb{P}(|Y - \mathbb{E}(Y)| > \|f\|_\infty \|\psi\|_\infty n |\Omega| \lambda) \leq 2 \exp \left(-\frac{\|f\|_\infty n |\Omega| \lambda^2}{2 + 4\lambda/3} \right).$$

Restricting $\lambda \leq 1$ completes the proof. \square

Remark 7. *If $\Omega = B(x, 2h)$ in Lemma 1, then under the same assumptions as in Lemma 1 we have*

$$(34) \quad \mathbb{P}(|Y - \mathbb{E}(Y)| \geq C \|\psi\|_\infty n h^d \lambda) \leq 2 \exp(-c n h^d \lambda^2),$$

for all $0 < \lambda \leq 1$, where $C, c > 0$ are constants depending only on $\|f\|_\infty$ and d .

3.2 Consistency of the 2-Laplacian

We now prove consistency for the graph 2-Laplacian.

Theorem 5. *Let $\delta > 0$ and let $\Omega_{\delta,n}$ denote the event that*

$$(35) \quad \left| \frac{1}{nh_n^d} d_n(x) - C_1 f(x) \right| \leq C\delta h_n,$$

and

$$(36) \quad \left| \frac{1}{nh_n^{d+2}} L_n^2 \varphi(x) - C_2 f^{-1} \operatorname{div}(f^2 \nabla \varphi) \right| \leq C \|\varphi\|_{C^3(B(x, 2h_n))} \delta,$$

hold for all $x \in X_n$ with $\operatorname{dist}(x, \mathcal{O}) > 2h_n$ and $\varphi \in C^3(\mathbb{R}^d)$. Then there exists $C, c > 0$ such that for $h_n \leq \delta \leq h_n^{-1}$ we have

$$(37) \quad \mathbb{P}[\Omega_{\delta,n}] \geq 1 - C \exp\left(-c\delta^2 nh_n^{d+2} + \log(n)\right).$$

Remark 8. We mention that $d_n(x)$ is a kernel density estimator for the density $f(x)$, and the error estimate (35) is standard in the density estimation literature [49].

Proof. By conditioning on the location of $x \in X_n$, we can assume without loss of generality that $x \in \mathbb{T}^d$ is a fixed (non-random) point. Write $\beta = \|\varphi\|_{C^3(B(x, 2h_n))}$. Let $\varphi \in C^3(\mathbb{R}^d)$ and set $p = D\varphi(x)$ and $A = D^2\varphi(x)$. Note that

$$(38) \quad \begin{aligned} L_n^2 \varphi(x) &= \sum_{i=1}^d p_i \sum_{y \in X_n} w_n(x, y)(y_i - x_i) + \frac{1}{2} \sum_{i,j=1}^d a_{ij} \sum_{y \in X_n} w_n(x, y)(y_i - x_i)(y_j - x_j) \\ &\quad + O(h^3 \beta d_n(x)). \end{aligned}$$

Let $\delta > 0$. By Lemma 1 and Remark 7, each of

$$\begin{aligned} \left| d_n(x) - n \int_{B(x, 2h_n)} \Phi\left(\frac{|x-y|}{h_n}\right) f(y) dy \right| &\geq C\delta n h_n^{d+1}, \\ \left| \sum_{y \in X_n} w_n(x, y)(y_i - x_i) - n \int_{B(x, 2h_n)} \Phi\left(\frac{|x-y|}{h_n}\right) (y_i - x_i) f(y) dy \right| &\geq C\delta n h_n^{d+2}, \end{aligned}$$

and

$$\begin{aligned} \left| \sum_{y \in X_n} w_n(x, y)(y_i - x_i)(y_j - x_j) - n \int_{B(x, 2h_n)} \Phi\left(\frac{|x-y|}{h_n}\right) (y_i - x_i)(y_j - x_j) f(y) dy \right| \\ \geq C\delta n h_n^{d+3}, \end{aligned}$$

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

$$(39) \quad \frac{1}{n h_n^{d+2}} L_n^2 \varphi(x) = \int_{B(0,2)} \Phi(|z|) \left(\frac{1}{h_n} p \cdot z + \frac{1}{2} z \cdot A z \right) f(x + z h_n) dz + O(\delta \beta)$$

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

$$\begin{aligned} \frac{1}{h_n} \int_{B(0,2)} \Phi(|z|) (p \cdot z) f(x + z h_n) dy &= \nabla f(x) \cdot \int_{B(0,2)} \Phi(|z|) (p \cdot z) z dz + O(\beta h_n) \\ &= \nabla f(x) \cdot \sum_{i=1}^d p_i \int_{B(0,2)} \Phi(|z|) z_i z dz + O(\beta h_n) \\ &= 2C_2 \nabla f(x) \cdot p + O(\beta h_n), \end{aligned}$$

and

$$\begin{aligned} \frac{1}{2} \int_{B(0,2)} \Phi(|z|) (z \cdot A z) f(x + z h_n) dz &= \frac{1}{2} f(x) \sum_{i,j=1}^d a_{ij} \int_{B(0,2)} \Phi(|z|) z_i z_j dz + O(h_n \beta) \\ &= C_2 f(x) \text{Trace}(A) + O(h_n \beta). \end{aligned}$$

Combining this with (39) we have that

$$(40) \quad \frac{1}{n h_n^{d+2}} L_n^2 \varphi(x) = C_2 f^{-1} \text{div}(f^2 \nabla \varphi) + O(\delta \beta),$$

holds with probability at least $1 - C \exp(-c\delta^2 n h_n^{d+2})$. The proof is completed by union bounding over all $x \in X_n$. \square

3.3 Consistency of the ∞ -Laplacian

We now prove consistency of the ∞ -Laplace operator. We define

$$(41) \quad H_n u(x) := \max_{y \in \mathbb{T}^d} w_n(x, y) (u(y) - u(x)) + \min_{y \in \mathbb{T}^d} w_n(x, y) (u(y) - u(x)),$$

Note that H_n is a non-local operator on the Torus, and does not depend on the realization of the random graph \mathcal{G}_n . Recalling the definition of the weights (13), we see that H_n depends only on the choice of length scale h_n in the problem.

We first recall a result from [23].

Lemma 2 (Lemma 1 in [23]). *Let $\varphi \in C^2(\mathbb{R}^d)$. Then*

$$(42) \quad |L_n^\infty \varphi(x) - H_n \varphi(x)| \leq C (\|\varphi\|_{C^1(B(x, 2h_n))} + h_n \|\varphi\|_{C^2(B(x, 2h_n))}) r_n^2 h_n^{-1},$$

where

$$(43) \quad r_n = \sup_{y \in \mathbb{T}^d} \text{dist}(y, \mathcal{X}_n).$$

Due to Lemma 2, we only require a consistency result for H_n .

Theorem 6. For any $x_0 \in \mathbb{T}^d \setminus \mathcal{O}$ and $\varphi \in C^3(\mathbb{R}^d)$ with $\nabla\varphi(x_0) \neq 0$

$$(44) \quad \lim_{\substack{n \rightarrow \infty \\ x \rightarrow x_0}} \frac{1}{h_n^2} H_n \varphi(x) = s_0^2 \Phi(s_0) \Delta_\infty \varphi(x_0),$$

where

$$\Delta_\infty \varphi = \frac{1}{|\nabla \varphi|^2} \sum_{i,j=1}^d \varphi_{x_i x_j} \varphi_{x_i} \varphi_{x_j}.$$

Proof. As in [23, Lemma 1] we define

$$B(x) = \max_{0 \leq s \leq 2} \left\{ \frac{1}{h_n} s \Phi(s) |\nabla \varphi(x)| - \frac{1}{2} s^2 \Phi(s) \Delta_\infty \varphi(x) \right\},$$

and

$$B'(x) = \max_{0 \leq s \leq 2} \left\{ \frac{1}{h_n} s \Phi(s) |\nabla \varphi(x)| + \frac{1}{2} s^2 \Phi(s) \Delta_\infty \varphi(x) \right\},$$

and we have

$$\frac{1}{h_n^2} H_n \varphi(x) = B'(x) - B(x) + O(\|\varphi\|_{C^3(B(x, 2h_n))} h_n).$$

Let $x_n \rightarrow x_0$ and for each n let $s_n, s'_n > 0$ be such that

$$B(x_n) = \frac{1}{h_n} s_n \Phi(s_n) |\nabla \varphi(x_n)| - \frac{1}{2} s_n^2 \Phi(s_n) \Delta_\infty \varphi(x_n),$$

and

$$B'(x_n) = \frac{1}{h_n} s'_n \Phi(s'_n) |\nabla \varphi(x_n)| + \frac{1}{2} (s'_n)^2 \Phi(s'_n) \Delta_\infty \varphi(x_n).$$

Then we have that

$$\frac{1}{h_n^2} H_n \varphi(x_n) \leq (s'_n)^2 \Phi(s'_n) \Delta_\infty \varphi(x_n) + O(\|\varphi\|_{C^3(B(x_n, 2h_n))} h_n),$$

and

$$\frac{1}{h_n^2} H_n \varphi(x_n) \geq s_n^2 \Phi(s_n) \Delta_\infty \varphi(x_n) + O(\|\varphi\|_{C^3(B(x_n, 2h_n))} h_n).$$

Since $s \mapsto s \Phi(s)$ has a unique maximum $s_0 \in (0, 2]$, we find that

$$s_n, s'_n \rightarrow s_0 \text{ as } n \rightarrow \infty.$$

It follows that

$$\lim_{n \rightarrow \infty} \frac{1}{h_n^2} H_n \varphi(x_n) = s_0^2 \Phi(s_0) \Delta_\infty \varphi(x_0).$$

□

4 Regularity

Here, we prove a discrete Hölder regularity result (Theorem 2) for the solutions u_n of (19). The proof is based on a well-known trick for establishing Hölder regularity for solutions of the unweighted p -Laplace equation

$$(45) \quad \operatorname{div}(|\nabla u|^{p-2} \nabla u) = 0$$

via the maximum principle. The trick is to notice that $v(x) = |x - y|^{\frac{p-d}{p-1}}$ is a solution of (45) away from $x = y$. When $p > d$, v is continuous and if we choose a large enough constant C so that

$$(46) \quad u(y) - Cv(x) \leq u(x) \leq u(y) + Cv(x)$$

for all boundary points x , then the maximum principle can be invoked to establish that (46) holds for all x , i.e.,

$$|u(y) - u(x)| \leq C|x - y|^{\frac{p-d}{p-1}}.$$

The function v is called a *barrier* in the PDE literature.

Adapting this to the graph setting is somewhat technical, since $v(x)$ is not an exact solution of the game theoretic p -Laplacian, due to random fluctuations and errors in the consistency results. The argument can be rescued by using the barrier $w(x) = |x - y|^\alpha$ for any $\alpha < (p - d)/(p - 1)$. This function is a strict supersolution of (45), which allows some room to account for the difference between the graph and continuum p -Laplacians. It is possible to show (see Lemma 3) that w is a supersolution of the game theoretic graph p -Laplacian (i.e., $L_n^p w(x) \leq 0$) for $|x - y| \geq ch_n$ for some $c > 0$ with high probability. The errors in the consistency results blow up as we approach the singularity at $x = y$, so w is not a global supersolution. In Lemma 4, we show how to modify w near $x = y$ to ensure the supersolution property holds globally. The modification relies extensively on the presence of the graph ∞ -Laplacian term.

For notational simplicity, we set

$$\Delta_2 \varphi = f^{-1} \operatorname{div}(f^2 \nabla \varphi).$$

We first require some elementary propositions.

Proposition 1. *For $p > d$ and $\alpha \in (0, 1)$ the function $v(x) = |x|^\alpha$ satisfies*

$$(47) \quad \Delta_2 v(x) \leq f(x) \alpha |x|^{\alpha-2} (d + \alpha - 2 + C|x|).$$

Proof. Notice that

$$\Delta_2 v = f \Delta v + 2 \nabla f \cdot \nabla v.$$

The proof is completed by computing

$$\Delta v = \alpha |x|^{\alpha-2} (d + \alpha - 2) \text{ and } \nabla v = \alpha |x|^{\alpha-2} x. \quad \square$$

Proposition 2. *Let $p > d$ and $\alpha \in (0, 1)$. For every $\delta > 0$ there exists $C \geq 2$ such that the function $v(x) = |x|^\alpha$ satisfies*

$$(48) \quad \frac{1}{h_n^2} H_n v(x) \leq (1 - \delta) s_0^2 \Phi(s_0) \alpha (\alpha - 1) |x|^{\alpha-2}$$

for all $|x| \geq Ch_n$.

Proof. Note that

$$\begin{aligned} H_n v(x) &= \max_{z \in B(0, 2)} \Phi(|z|) (|x + h_n z|^\alpha - |x|^\alpha) + \min_{z \in B(0, 2)} \Phi(|z|) (|x + h_n z|^\alpha - |x|^\alpha) \\ &= \max_{0 \leq s \leq 2} \Phi(s) ((|x| + s h_n)^\alpha - |x|^\alpha) + \min_{0 \leq s \leq 2} \Phi(s) ((|x| - s h_n)^\alpha - |x|^\alpha). \end{aligned}$$

Let $s_n \geq 0$ such that

$$A_n := \Phi(s_n) ((|x| + s_n h_n)^\alpha - |x|^\alpha) = \max_{0 \leq s \leq 2} \Phi(s) ((|x| + s h_n)^\alpha - |x|^\alpha).$$

Then we have that

$$H_n v(x) \leq \Phi(s_n) ((|x| + s_n h_n)^\alpha + (|x| - s_n h_n)^\alpha - 2|x|^\alpha).$$

By Taylor expanding the right-hand side we have

$$(49) \quad H_n v(x) \leq s_n^2 \Phi(s_n) \alpha (\alpha - 1) |x|^{\alpha-2} h_n^2.$$

Note that

$$\begin{aligned} A_n &= \Phi(s_n) ((|x| + s_n h_n)^\alpha - |x|^\alpha) \\ &\leq \Phi(s_n) (|x|^\alpha + \alpha s_n h_n |x|^{\alpha-1} - |x|^\alpha) \\ &= s_n \Phi(s_n) \alpha |x|^{\alpha-1} h_n. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} A_n &\geq \Phi(s_0) ((|x| + s_0 h_n)^\alpha - |x|^\alpha) \\ &\geq \Phi(s_0) \left(|x|^\alpha + \alpha s_0 h_n |x|^{\alpha-1} + \frac{\alpha(\alpha-1)}{2} s_0^2 h_n^2 |x|^{\alpha-2} - |x|^\alpha \right) \\ &= \frac{1}{2} s_0 \Phi(s_0) \alpha |x|^{\alpha-1} h_n (2 + (\alpha-1) s_0 h_n |x|^{-1}). \end{aligned}$$

Combining these two inequalities we find that

$$(50) \quad s_n \Phi(s_n) \geq \left(1 - \frac{1}{2}(1 - \alpha) s_0 h_n |x|^{-1}\right) s_0 \Phi(s_0).$$

By (12) we have

$$\theta(s_n - s_0)^2 \leq s_0\Phi(s_0) - s_n\Phi(s_n) \leq \frac{1}{2}(1 - \alpha)s_0^2\Phi(s_0)h_n|x|^{-1}.$$

Therefore

$$s_n \geq \left(1 - \theta^{-1/2}(1 - \alpha)^{1/2}h_n^{1/2}|x|^{-1/2}\right)s_0.$$

Combining this with (50), we see that for every $\delta > 0$ there exists $C > 0$ such that

$$s_n^2\Phi(s_n) \geq (1 - \delta)s_0^2\Phi(s_0)$$

for $|x| \geq Ch_n$. The proof is completed by inserting this into (49) and noting that $\alpha - 1 < 0$. \square

We now establish that our barrier function $|x - y|^\alpha$ is a supersolution away from $x = y$.

Lemma 3. *Let $p > d$, $\sigma \in [0, 1]$, $\alpha < (p-d)/(p-1)$, and for $y \in \mathbb{T}^d$, and set $v_y(x) = |x - y|^\alpha$. Then there exists $C, c, \delta, r > 0$, depending only on d, p , and α , so that*

$$\mathbb{P}[\forall y \in \mathcal{X}_n, x \in X_n \cap B(y, r) \setminus B^0(y, ch_n^\sigma), L_n^p v_y(x) \leq 0] \geq 1 - \exp(-\delta n h_n^q + C \log(n)),$$

where $q = \max\{d + 2 + 2\sigma, 3d/2\}$. Furthermore, if $\sigma < 1$ then the Lemma holds for any $c > 0$, provided h_n is sufficiently small, and C, δ , and r now additionally depend on c .

Proof. Contrary to Theorem 5 we may have $\text{dist}(x, \mathcal{O}) \leq 2h_n$, and must account for this. Notice that

$$L_n^2 v_y(x) = \sum_{z \in X_n} w_n(x, z)(v_y(z) - v_y(x)) + \sum_{z \in \mathcal{O}} w_n(x, z)(v_y(z) - v_y(x)),$$

and

$$d_n(x) = \sum_{z \in X_n} w_n(x, z) + \sum_{z \in \mathcal{O}} w_n(x, z).$$

The argument from Theorem 5 applies to the summations over X_n in the equations above, and the summations over \mathcal{O} are bounded by a constant. Thus, assuming $nh_n^{d+2} \geq 1$ and $|x - y| \geq \max\{h_n^\sigma, 3h_n\}$, Theorem 5 and Proposition 1 yield

$$\begin{aligned} \frac{L_n^2 v_y(x)}{d_n(x)h_n^2} &\leq \frac{C_2 \Delta_2 v_y(x) + C(|x - y|^{\alpha-3} \delta h_n^\sigma + n^{-1} h_n^{-d-2})}{C_1 f(x) - C(\delta h_n^{1+\sigma} + n^{-1} h_n^{-d})} \\ &\leq \frac{C_2 \alpha f(x) |x - y|^{\alpha-2} (d + \alpha - 2 + C|x - y| + C\delta)}{C_1 f(x) - C\delta h_n^{1+\sigma}} \\ (51) \quad &\leq \frac{C_2}{C_1} \alpha |x - y|^{\alpha-2} (d + \alpha - 2 + C|x - y| + C\delta). \end{aligned}$$

holds for all $y \in \mathcal{X}_n$ and $x \in X_n$ with $|x - y| \geq \max\{h_n^\sigma, 3h_n\}$ with probability at least

$$(52) \quad 1 - C \exp\left(-\delta^2 n h_n^{d+2+2\sigma} + \log(n)\right).$$

For the rest of the proof we assume that (51) holds.

Let $\delta' > 0$. By Proposition 2 there exists $C_{\delta'}$ such that

$$\frac{1}{h_n^2} H_n v_y(x) \leq (1 - \delta') s_0^2 \Phi(s_0) \alpha(\alpha - 1) |x - y|^{\alpha-2}$$

for all $|x - y| \geq C_{\delta'} h_n$. By Lemma 2 we have

$$L_n^\infty v_y(x) \leq H_n v_y(x) + C(|x - y|^{\alpha-1} + h_n |x - y|^{\alpha-2}) r_n^2 h_n^{-1}$$

whenever $|x - y| \geq 3h_n$. Therefore

$$\frac{1}{h_n^2} L_n^\infty v_y(x) \leq \alpha |x - y|^{\alpha-2} s_0^2 \Phi(s_0) ((1 - \delta')(\alpha - 1) + C |x - y| r_n^2 h_n^{-3})$$

whenever $|x - y| \geq C_{\delta'} h_n$. Combining this with (51) gives

$$h_n^{-2} L_n^p v_y(x) \leq C |x - y|^{\alpha-2} (\alpha(p - 1) + d - p + C |x - y| + C |x - y| r_n^2 h_n^{-3} + C\delta + C\delta')$$

for all $y \in \mathcal{X}_n$ and $x \in X_n$ with $|x - y| \geq \max\{h_n^\sigma, C_{\delta'} h_n\}$.

We now bound r_n . Let $t > 0$ and partition \mathbb{T}^d into boxes $R_1, R_2, R_3, \dots, R_M$ of side length t/\sqrt{d} . If $r_n > t$ then at least one box R_i contains no points from X_n . It follows that

$$\begin{aligned} \mathbb{P}[r_n > t] &\leq \sum_{i=1}^M \mathbb{P}[X_n \cap R_i = \emptyset] \\ &\leq M(1 - Ct^d)^n \\ &\leq C \exp\left(n \log(1 - Ct^d) - d \log(t)\right) \\ &\leq C \exp\left(-Cnt^d - d \log(t)\right). \end{aligned}$$

Since we assume that $h_n \geq n^{-1/d}$ we have that

$$\mathbb{P}[r_n^2 > h_n^3] \leq C \exp\left(-Cn h_n^{3d/2} + 4 \log(n)\right).$$

The proof is completed by choosing $\delta, \delta' > 0$ sufficiently small and noting that by assumption we have $\alpha(p - 1) + d - p < 0$. \square

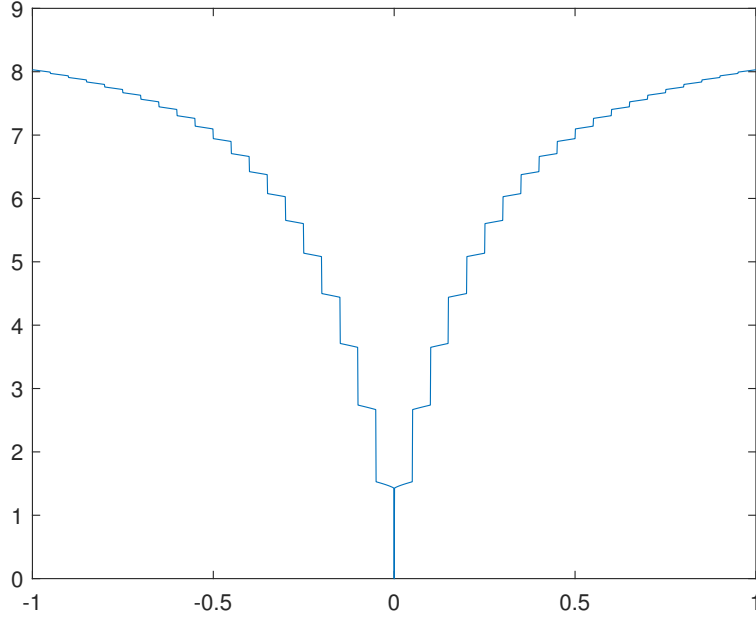


Figure 2: A depiction of the local barrier, defined in (53), used to complete the proof of Hölder regularity.

In order to use the barrier function technique to prove a global Hölder regularity result, we need a barrier that is a supersolution globally, including within the local $O(h_n)$ neighborhood of $x = y$ where the barrier from Lemma 3 fails to be a supersolution. In Lemma 4 we show how to construct a barrier in this local neighborhood using a construction that heavily exploits the min/max structure of the ∞ -Laplace term. Before giving the proof and the explicit form for the barrier, let us describe the idea heuristically. The new local barrier, defined in Equation (53) below and depicted in Figure 2, consists of sharp jump discontinuities that decay in size rapidly away from the origin. The spacing between jumps is $h_n/2$ to ensure that in the definition of the ∞ -Laplacian, the min term will take a large negative jump, while the max term (and the 2-Laplacian term) will take a much smaller positive jump, resulting in a large negative value for $L_n^p v_y(x)$, even at points arbitrarily close to the origin. We note that this barrier does not depend on the $p > d$ assumption and works for arbitrary p . The sharp decay in the jump size away from the origin prevents the barrier from being a supersolution outside of this $O(h_n)$ neighborhood when $p < d$.

We now give the statement and proof of Lemma 4.

Lemma 4. *Let $p > 2$ and $0 < \alpha < 1$. For $M > 0$ and $y \in \mathbb{T}^d$ define*

$$(53) \quad v_y(x) = |x - y|^\alpha + M h_n^\alpha \sum_{k=1}^{\infty} \beta^k 1_{\{2|x-y| > (k-1)h_n\}},$$

where

$$\beta = \frac{\lambda(p-2)}{8(m\lambda(p-2)+1)} \quad \text{and} \quad m = \max_{0 \leq s \leq 2} \Phi(s).$$

Then for every $c > 0$ there exists $M > 0$ such that

$$\mathbb{P}[\forall y \in \mathcal{X}_n, x \in X_n \cap B(y, (c+4)h_n) \setminus \{y\}, L_n^p v_y(x) \leq 0] \geq 1 - \exp\left(-Cnh_n^d + 2\log(n)\right).$$

Proof. Let $c > 0$ and let Ω denote the event that for every $y \in \mathcal{X}_n$ and every $x \in B(y, (c+4)h_n) \cap X_n$ with $|x-y| > h_n$, the set

$$B_x := B(x, h_n) \cap B(y, |x-y| - h_n/2)$$

has a nonempty intersection with X_n . There exists $a > 0$ such that $|B_x| \geq ah_n^d$ for all $|x| > h_n$. Therefore $\mathbb{P}[\Omega] \geq 1 - n^2 e^{-Cnh_n^d}$. For the rest of the proof, we assume Ω occurs.

Let $y \in \mathcal{X}_n$, $x \in B(y, (c+4)h_n) \cap X_n$ and let $k \in \mathbb{N}$ such that $(k-1)h_n < 2|x-y| \leq kh_n$. For $z \in B(x, 2h_n)$ we have

$$\begin{aligned} v_y(z) - v_y(x) &\leq Mh^\alpha(\beta^{k+4} + \beta^{k+3} + \beta^{k+2} + \beta^{k+1}) + (|x-y| + 2h_n)^\alpha - |x-y|^\alpha \\ &\leq 4M\beta^{k+1}h_n^\alpha + ((c+4)h_n + 2h_n)^\alpha \\ &\leq 4M\beta^{k+1}h_n^\alpha + (c+6)^\alpha h_n^\alpha. \end{aligned}$$

Therefore

$$\frac{1}{d_n(x)} L_n^2 v_y(x) \leq 4M\beta^{k+1}h_n^\alpha + (c+6)^\alpha h_n^\alpha,$$

and

$$\max_{z \in X_n} w_n(x, z)(v_y(z) - v_y(x)) \leq 4Mm\beta^{k+1}h_n^\alpha + m(c+6)^\alpha h_n^\alpha.$$

If $k = 1, 2$, then $y \in B(x, h_n)$ and since $v_y(y) = 0$ we have

$$\min_{z \in X_n} w_n(x, z)(v_y(z) - v_y(x)) \leq -M\beta^k h_n^\alpha,$$

If $k \geq 3$, then $|x-y| > h_n$. Since $B_x \cap X_n$ is nonempty, there exists $z \in B(x, h_n) \cap X_n$ such that $2|z-y| \leq (k-1)h_n$, and hence

$$\min_{z \in X_n} w_n(x, z)(v_y(z) - v_y(x)) \leq -M\beta^k h_n^\alpha.$$

Combining the above observations we have

$$\begin{aligned} h_n^{-\alpha} L_n^p v_y(x) &\leq 4M\beta^{k+1} + (c+6)^\alpha + \lambda(p-2) \left(4Mm\beta^{k+1} + m(c+6)^\alpha - M\beta^k \right) \\ &= (c+6)^\alpha (1 + 2\lambda(p-2)) - \frac{1}{2}\lambda(p-2)\beta^k M, \end{aligned}$$

for $x \in B(y, (c+4)h_n) \cap X_n$ with $y \neq x$. Now we can simply choose $M > 0$ large enough so that $L_n^p v_y(x) \leq 0$ for all $k \leq 2c+1$. \square

We are now ready to give the proof of Theorem 2. The proof involves patching together the barrier functions provided in Lemmas 3 and 4, and using the global barrier to establish Hölder regularity.

Proof of Theorem 2. Choose $r, c, \delta > 0$ so that the conclusions of Lemma 3 hold for $\sigma = 1$. We may assume without loss of generality that c is a positive integer. Define β as in Lemma 4 and let $M > 0$ so that the conclusions of Lemma 4 hold with the aforementioned value of c . For the rest of the theorem, we assume the conclusions of Lemmas 3 and 4 hold, which is an event with probability at least $1 - \exp(-\delta n h_n^q + C \log(n))$, where $q = \max\{d + 4, 3d/2\}$.

The proof is now split into three steps.

1. Let $y \in \mathcal{X}_n$, define $k_0 = 2c + 5$ and set

$$v_1(x) = |x - y|^\alpha + M h_n^\alpha \sum_{k=1}^{\infty} \beta^k 1_{\{2|x-y| > (k-1)h_n\}},$$

$$v_2(x) = |x - y|^\alpha + \frac{M \beta h_n^\alpha}{1 - \beta} (1 - \beta^{k_0}),$$

and

$$v_y(x) = \min \{v_1(x), v_2(x)\}.$$

We claim that $L_n^p v_y(x) \leq 0$ for all $0 < |x - y| \leq r$. To see this, notice that for $|x - y| \leq (c + 2)h_n$ we have $v_y(x) = v_1(x)$ and for $|x - y| \geq (c + 2)h_n$ we have $v_y(x) = v_2(x)$. By Lemmas 3 and 4 we see that $L_n^p v_1(x) \leq 0$ whenever $|x - y| \leq (c + 4)h_n$ and $L_n^p v_2(x) \leq 0$ whenever $|x - y| \geq ch_n$. It follows that $L_n^p v_y(x) \leq 0$ for all $x \neq y$ such that $r \geq |x - y| \geq (c + 4)h_n$ or $|x - y| \leq ch_n$. If $(c + 2)h_n \leq |x - y| \leq (c + 4)h_n$ then $v_y(x) = v_2(x)$ and we have $L_n^p v_y(x) \leq L_n^p v_2(x) \leq 0$ because $v \leq v_2$ in $B(x, 2h_n)$. Likewise, if $ch_n \leq |x - y| \leq (c + 2)h_n$ then $L_n^p v_y(x) \leq L_n^p v_1(x) \leq 0$. This establishes the claim.

2. Let $y \in \mathcal{O}$. Let $C > 0$ be large enough so that for all $x, y \in \mathcal{O}$

$$C v_y(x) \geq \max_{\mathcal{O}} g - \min_{\mathcal{O}} g.$$

It follows that

$$g(y) - C v_y(x) \leq u_n(x) \leq g(y) + C v_y(x)$$

holds for all $x \in \mathcal{O}$. By the maximum principle (Theorem 3) and the fact that $u_n(y) = g(y)$ we have that

$$(54) \quad u_n(y) - C v_y \leq u_n \leq u_n(y) + C v_y$$

for all $y \in \mathcal{O}$.

3. Let $y \in X_n$. We claim that

$$u_n(y) - C v_y(z) \leq u_n(z) \leq u_n(y) + C v_y(z)$$

holds for all $z \in \mathcal{O}$. Indeed, recalling (54) we have

$$u_n(y) \geq u_n(z) - Cv_z(y) = u_n(z) - Cv_y(z),$$

and

$$u_n(y) \leq u_n(z) + Cv_z(y) = u_n(z) + Cv_y(z),$$

which establishes the claim. By the maximum principle (Theorem 3) we have that

$$(55) \quad u_n(y) - Cv_y \leq u_n \leq u_n(y) + Cv_y$$

for all $y \in X_n$.

4. Notice that

$$(56) \quad |x - y|^\alpha \leq v_y(x) \leq |x - y|^\alpha + \frac{M\beta h_n^\alpha}{1 - \beta}.$$

Combining this with (54) and (55) we have

$$(57) \quad u_n(y) - C(|x - y|^\alpha + h_n^\alpha) \leq u_n(x) \leq u_n(y) + C(|x - y|^\alpha + h_n^\alpha),$$

for all $x, y \in \mathcal{X}_n$, which completes the proof. \square

5 Convergence proof

In this section we prove Theorem 1. We first recall the notion of viscosity solution for

$$(58) \quad F(\nabla^2 u, \nabla u, x) = 0 \quad \text{in } \Omega,$$

where $\Omega \subset \mathbb{R}^d$ is open.

Definition 1. We say that $u \in C(\Omega)$ is a viscosity subsolution of (58) if for every $x_0 \in \Omega$ and $\varphi \in C^\infty(\Omega)$ such that $u - \varphi$ has a local maximum at x_0 we have

$$F(\nabla^2 \varphi(x_0), \nabla \varphi(x_0), x_0) \leq 0.$$

We say that $u \in C(\Omega)$ is a viscosity supersolution of (58) if for every $x_0 \in \Omega$ and $\varphi \in C^\infty(\Omega)$ such that $u - \varphi$ has a local minimum at x_0 we have

$$F(\nabla^2 \varphi(x_0), \nabla \varphi(x_0), x_0) \geq 0.$$

We say that $u \in C(\Omega)$ is a viscosity solution of (58) if u is both a viscosity sub- and supersolution of (58).

Let $\pi : \mathbb{R}^d \rightarrow \mathbb{T}^d$ be the projection operator.

Definition 2. A function $u \in C(\mathbb{T}^d)$ is a viscosity solution of (22) if $u = g$ on \mathcal{O} and $v(x) := u(\pi(x))$ is a viscosity solution of (58) with

$$F_p(X, q, x) := -|q|^{p-2}(\text{Trace}(X) + 2\nabla \log f(x) \cdot q + (p-2)|q|^{-2}q \cdot Xq),$$

and $\Omega := \pi^{-1}(\mathbb{T}^d \setminus \mathcal{O})$.

We first verify uniqueness of viscosity solutions of (22).

Theorem 7. Assume $p > d \geq 2$ and $f \in C^2(\mathbb{T}^d)$ is positive. Then there exists a unique viscosity solution of (22).

The proof of Theorem 7 follows [50] closely, with some minor adjustments to handle the weighted p -Laplacian, and some simplifications due to our assumption that $p > d \geq 2$. In particular, the proof shows that weak and viscosity solutions of (22) coincide.

Proof. Existence follows from the proof of Theorem 1. We only need to prove uniqueness here.

Let $u \in C(\mathbb{T}^d)$ be a viscosity solution of (22). Extending u to \mathbb{R}^d by setting $u(x) = u(\pi(x))$, we have that $u \in C(\mathbb{R}^d)$ is a \mathbb{Z}^d -periodic viscosity solution of

$$(59) \quad \left. \begin{aligned} F_p(\nabla^2 u, \nabla u, x) &= 0 & \text{in } \Omega \\ u &= g & \text{on } \partial\Omega, \end{aligned} \right\}$$

where $\Omega = \pi^{-1}(\mathbb{T}^d \setminus \mathcal{O})$, $f(x) = f(\pi(x))$ and $g(x) = g(\pi(x))$. For $\varepsilon \geq 0$, let $u_\varepsilon \in W_{per}^{1,p}(\mathbb{R}^d)$ be the unique \mathbb{Z}^d -periodic weak solution (defined via integration by parts) of

$$(60) \quad \left. \begin{aligned} -\text{div}(f^2|\nabla u_\varepsilon|^{p-2}\nabla u_\varepsilon) &= \varepsilon & \text{in } \Omega \\ u_\varepsilon &= g & \text{on } \partial\Omega. \end{aligned} \right\}$$

The solutions u_ε can be constructed by the Calculus of Variations, for instance. By regularity theory for degenerate elliptic equations [33, 34], we have $u_\varepsilon \in C_{loc}^{1,\gamma}(\Omega)$, and since $p > d$ we have $u_\varepsilon \in C^{0,1-d/p}(\mathbb{R}^d)$ by Morrey's inequality.

We will show that $u = u_0$. The proof is split into 4 steps.

1. First, we claim that for $\varepsilon > 0$, u_ε is a viscosity supersolution of

$$(61) \quad f(x)^2 F_p(\nabla^2 u_\varepsilon, \nabla u_\varepsilon, x) = \varepsilon \quad \text{in } \Omega.$$

To see this, let $x_0 \in \Omega$ and $\varphi \in C^\infty(\Omega)$ such that $u_\varepsilon - \varphi$ has a local minimum at x_0 . We need to show that

$$-\text{div}(f^2|\nabla \varphi|^{p-2}\nabla \varphi) \Big|_{x=x_0} \geq \varepsilon.$$

Assume to the contrary that

$$-\text{div}(f^2|\nabla \varphi|^{p-2}\nabla \varphi) \Big|_{x=x_0} < \varepsilon.$$

For $\delta > 0$ define

$$\psi(x) = \varphi(x) + \delta^2 + u_\varepsilon(x_0) - \varphi(x_0) - \delta|x - x_0|^4.$$

For sufficiently small $r, \delta > 0$ we have $u_\varepsilon(x) \geq \psi(x)$ for all $x \in \partial B(x_0, r)$, and

$$(62) \quad -\operatorname{div}(f^2|\nabla\psi|^{p-2}\nabla\psi) \leq \varepsilon \quad \text{in } B(x_0, r).$$

The comparison principle for weak solutions of the p -Laplace equation yields $\psi \leq u_\varepsilon$ in $B(x_0, r)$, which is a contradiction to the fact that $\psi(x_0) > u_\varepsilon(x_0)$. This establishes the claim.

2. We now show that $u \leq u_\varepsilon$ for $\varepsilon > 0$. Fix $\varepsilon > 0$ and for $\delta > 0$ define

$$\Omega_\delta = \{x \in \Omega : \operatorname{dist}(x, \partial\Omega) > \delta\}.$$

Let $M > 0$ such that $|\nabla u_\varepsilon| \leq M$ on Ω_δ and define the truncated operator

$$\bar{F}_p(X, q, x) := -\min\{M, |q|\}^{p-2}(\operatorname{Trace}(X) + 2\nabla \log f(x) \cdot q + (p-2)|q|^{-2}q \cdot Xq).$$

Then u is clearly a viscosity solution of $\bar{F}_p = 0$ in Ω , and u_ε is a viscosity solution of $f^2\bar{F}_p \geq \varepsilon$ in Ω_δ . Since $f \geq \theta > 0$ for some $\theta > 0$, and \bar{F}_p satisfies

$$\bar{F}_p(y, q, Y) - \bar{F}_p(x, q, X) \leq 2M^{p-2}[\nabla f]_{0,1}|q||x - y|,$$

for all $q \in \mathbb{R}^d$, $x, y \in \Omega_\delta$ and symmetric matrices $X \leq Y$, we can invoke the comparison principle for viscosity solutions [35] to obtain

$$\max_{\partial\Omega_\delta}(u - u_\varepsilon) = \max_{\Omega_\delta}(u - u_\varepsilon).$$

Since u, u_ε are uniformly continuous on \mathbb{R}^d , we can send $\delta \rightarrow 0^+$ to find that $u \leq u_\varepsilon$ on \mathbb{R}^d .

3. We now send $\varepsilon \rightarrow 0^+$ to show that $u \leq u_0$. Using $\varphi = \max\{u_\varepsilon - u_0 - \delta, 0\}$ as a test function in the definition of weak solution of (60) for u_ε and u_0 and subtracting, we have

$$\int_{A_\delta} f^2|\nabla u_\varepsilon - \nabla u_0|^p dx \leq C \int_{A_\delta} f^2(|\nabla u_\varepsilon|^{p-2}\nabla u_\varepsilon - |\nabla u_0|^{p-2}\nabla u_0) \cdot (\nabla u_\varepsilon - \nabla u_0) dx \leq C\varepsilon,$$

where $A_\delta = \{x \in (0, 1)^d : u_\varepsilon(x) - u_0(x) > \delta\}$ and $\delta > 0$. Sending $\delta \rightarrow 0^+$ we find that

$$\int_{\{u_\varepsilon > u\}} f^2|\nabla u_\varepsilon - \nabla u_0|^p dx \leq C\varepsilon.$$

A similar argument gives

$$\int_{\{u_\varepsilon < u\}} f^2|\nabla u_\varepsilon - \nabla u_0|^p dx \leq C\varepsilon.$$

By Morrey's inequality $u_\varepsilon \rightarrow u_0$ uniformly on \mathbb{R}^d as $\varepsilon \rightarrow 0^+$. Therefore $u \leq u_0$.

4. To see that $u \geq u_0$, we apply the argument above to $v := -u$ to find that $v \leq -u_0$, or $u \geq u_0$. Therefore $u = u_0$, which completes the proof. \square

We now give the proof of our main result.

Proof of Theorem 1. The proof is split into several steps.

1. Since

$$\lim_{n \rightarrow \infty} \frac{nh_n^q}{\log(n)} = \infty,$$

we can apply the Borel-Cantelli Lemma and Theorems 2, 5, 6, and Lemma 2 to show that with probability one

$$(63) \quad \lim_{\substack{n \rightarrow \infty \\ x \rightarrow x_0}} \frac{1}{h_n^2} L_n^p \varphi(x) = \frac{C_2}{C_1} f^{-2} |\nabla \varphi|^{2-p} \operatorname{div}(f^2 |\nabla \varphi|^{p-2} \nabla \varphi) \Big|_{x=x_0}$$

for all $x_0 \in \mathbb{T}^d \setminus \mathcal{O}$ and $\varphi \in C^\infty(\mathbb{R}^d)$ with $\nabla \varphi(x_0) \neq 0$, and

$$(64) \quad |u_n(x) - u_n(y)| \leq C(|x - y|^{1-d/p} + h_n^{1-d/p})$$

for n sufficiently large. For the remainder of the proof, we fix a realization in this probability one event.

2. Let $j_n : \mathbb{T}^d \rightarrow \mathcal{X}_n$ be the closest point projection, which satisfies

$$|x - j_n(x)| = \min_{y \in \mathcal{X}_n} |x - y|.$$

Define $v_n : \mathbb{T}^d \rightarrow \mathbb{R}$ by $v_n(x) = u_n(j_n(x))$. Since $|x - j_n(x)| \leq r_n$, where r_n is defined in (43), it follows from (64) that for any $x, y \in \mathbb{T}^d$

$$\begin{aligned} |v_n(x) - v_n(y)| &\leq C(|j_n(x) - j_n(y)|^{1-d/p} + h_n^{1-d/p}) \\ &= C(|j_n(x) - x + x - y + y - j_n(y)|^{1-d/p} + h_n^{1-d/p}) \\ &\leq C(|x - y|^{1-d/p} + 2r_n^{1-d/p} + h_n^{1-d/p}). \end{aligned}$$

Since $r_n, h_n \rightarrow 0$ as $n \rightarrow \infty$, we can use the Arzelà-Ascoli Theorem (see the appendix in [51]) to extract a subsequence, denote again by v_n , and a Hölder continuous function $u \in C^{0,1-d/p}(\mathbb{T}^d)$ such that $v_n \rightarrow u$ uniformly on \mathbb{T}^d as $n \rightarrow \infty$. Since $u_n(x) = v_n(x)$ for all $x \in \mathcal{X}_n$ we have

$$(65) \quad \lim_{n \rightarrow \infty} \max_{x \in \mathcal{X}_n} |u_n(x) - u(x)| = 0.$$

We claim that u is the unique viscosity solution of (22), which will complete the proof.

3. We first show that u is a viscosity subsolution of (22). Let $x_0 \in \mathbb{T}^d \setminus \mathcal{O}$ and $\varphi \in C^\infty(\mathbb{R}^d)$ such that $u - \varphi$ has a strict global maximum at the point x_0 and $\nabla \varphi(x_0) \neq 0$. We need to show that

$$\operatorname{div}(f^2 |\nabla \varphi|^{p-2} \nabla \varphi) \Big|_{x=x_0} \geq 0.$$

By (65) there exists a sequence of points $x_n \in \mathcal{X}_n$ such that $u_n - \varphi$ attains its global maximum at x_n and $x_n \rightarrow x_0$ as $n \rightarrow \infty$. Therefore

$$u_n(x_n) - u_n(x) \geq \varphi(x_n) - \varphi(x) \quad \text{for all } x \in \mathcal{X}_n.$$

Since $x_0 \notin \mathcal{O}$, we have that $x_n \notin \mathcal{O}$ for n sufficiently large and so $L_n^p u_n(x_n) \leq L_n^p \varphi(x_n)$. By (63) we have

$$0 = \lim_{n \rightarrow \infty} \frac{1}{h_n^2} L_n^p u_n(x_n) \leq \lim_{n \rightarrow \infty} \frac{1}{h_n^2} L_n^p \varphi(x_n) = \frac{C_2}{C_1} f^{-2} |\nabla \varphi|^{2-p} \operatorname{div}(f^2 |\nabla \varphi|^{p-2} \nabla \varphi)|_{x=x_0}.$$

Thus u is a viscosity subsolution of (22).

4. To verify the supersolution property, set $v_n = -u_n$ and note that $L_n^p v_n = -L_n^p u_n = 0$ and $v_n \rightarrow -u$ uniformly as $n \rightarrow \infty$. The subsolution argument above shows that $-u$ is a viscosity subsolution of (22), and hence u is a viscosity supersolution. This completes the proof. \square

Acknowledgments

The author gratefully acknowledges the support of NSF-DMS grant 1713691. The author is also grateful to the anonymous referees, whose suggestions have greatly improved the paper.

References

- [1] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. *Semi-supervised learning*. MIT, 2006.
- [2] Xiaojin Zhu, Zoubin Ghahramani, John Lafferty, et al. Semi-supervised learning using gaussian fields and harmonic functions. In *International Conference on Machine Learning*, volume 3, pages 912–919, 2003.
- [3] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning from labeled and unlabeled data on a directed graph. In *Proceedings of the 22nd International Conference on Machine Learning*, pages 1036–1043. ACM, 2005.
- [4] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. *Advances in Neural Information Processing Systems*, 16(16):321–328, 2004.
- [5] Dengyong Zhou, Jason Weston, Arthur Gretton, Olivier Bousquet, and Bernhard Schölkopf. Ranking on data manifolds. *Advances in Neural Information Processing Systems*, 16:169–176, 2004.

- [6] Rie Kubota Ando and Tong Zhang. Learning on graph with laplacian regularization. *Advances in Neural Information Processing Systems*, 19:25, 2007.
- [7] Jingrui He, Mingjing Li, Hong-Jiang Zhang, Hanghang Tong, and Changshui Zhang. Manifold-ranking based image retrieval. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 9–16. ACM, 2004.
- [8] Jingrui He, Mingjing Li, H-J Zhang, Hanghang Tong, and Changshui Zhang. Generalized manifold-ranking-based image retrieval. *IEEE Transactions on image processing*, 15(10):3170–3177, 2006.
- [9] Yang Wang, Muhammad Aamir Cheema, Xuemin Lin, and Qing Zhang. Multi-manifold ranking: Using multiple features for better image retrieval. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 449–460. Springer, 2013.
- [10] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 3166–3173, 2013.
- [11] Xueyuan Zhou, Mikhail Belkin, and Nathan Srebro. An iterated graph laplacian approach for ranking on manifolds. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 877–885. ACM, 2011.
- [12] Bin Xu, Jiajun Bu, Chun Chen, Deng Cai, Xiaofei He, Wei Liu, and Jiebo Luo. Efficient manifold ranking for image retrieval. In *Proceedings of the 34th international ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 525–534. ACM, 2011.
- [13] Boaz Nadler, Nathan Srebro, and Xueyuan Zhou. Semi-supervised learning with the graph Laplacian: The limit of infinite unlabelled data. In *Neural Information Processing Systems (NIPS)*, 2009.
- [14] Ahmed El Alaoui, Xiang Cheng, Aaditya Ramdas, Martin J Wainwright, and Michael I Jordan. Asymptotic behavior of lp-based Laplacian regularization in semi-supervised learning. In *29th Annual Conference on Learning Theory*, pages 879–906, 2016.
- [15] Nick Bridle and Xiaojin Zhu. p-voltages: Laplacian regularization for semi-supervised learning on high-dimensional data. In *Eleventh Workshop on Mining and Learning with Graphs (MLG2013)*, 2013.
- [16] Morteza Alamgir and Ulrike V Luxburg. Phase transition in the family of p-resistances. In *Advances in Neural Information Processing Systems*, pages 379–387, 2011.

- [17] Rasmus Kyng, Anup Rao, Sushant Sachdeva, and Daniel A Spielman. Algorithms for lipschitz learning on graphs. In *Proceedings of The 28th Conference on Learning Theory*, pages 1190–1223, 2015.
- [18] Ulrike von Luxburg and Olivier Bousquet. Distance-based classification with lipschitz functions. *Journal of Machine Learning Research*, 5(Jun):669–695, 2004.
- [19] Gunnar Aronsson, Michael Crandall, and Petri Juutinen. A tour of the theory of absolutely minimizing functions. *Bulletin of the American Mathematical Society*, 41(4):439–505, 2004.
- [20] Peter Lindqvist. *Notes on the p -Laplace equation*. 2017.
- [21] Yuval Peres, Oded Schramm, Scott Sheffield, and David Wilson. Tug-of-war and the infinity laplacian. *Journal of the American Mathematical Society*, 22(1):167–210, 2009.
- [22] Marta Lewicka and Juan J Manfredi. Game theoretical methods in pdes. *Bollettino dell’Unione Matematica Italiana*, 7(3):211–216, 2014.
- [23] Jeff Calder. Consistency of Lipschitz learning with infinite unlabeled and finite labeled data. *arXiv:1710.10364*, 2017.
- [24] Dejan Slepčev and Matthew Thorpe. Analysis of p -laplacian regularization in semi-supervised learning. *arXiv preprint arXiv:1707.06213*, 2017.
- [25] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [26] Matthias Hein and Jean-Yves Audibert. Intrinsic dimensionality estimation of submanifolds in \mathbb{R}^d . In *Proceedings of the 22nd International Conference on Machine learning*, pages 289–296. ACM, 2005.
- [27] Jose A Costa and Alfred O Hero. Determining intrinsic dimension and entropy of high-dimensional shape spaces. In *Statistics and Analysis of Shapes*, pages 231–252. Springer, 2006.
- [28] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.
- [29] Daniel A Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 81–90. ACM, 2004.

- [30] Michael B Cohen, Rasmus Kyng, Gary L Miller, Jakub W Pachocki, Richard Peng, Anup B Rao, and Shen Chen Xu. Solving SDD linear systems in nearly $m \log 1/2 n$ time. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 343–352. ACM, 2014.
- [31] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [32] Juan J Manfredi, Adam M Oberman, and Alexander P Sviridov. Nonlinear elliptic partial differential equations and p-harmonic functions on graphs. *Differential Integral Equations*, 28(1–2):79–102, 2015.
- [33] Peter Tolksdorf. Regularity for a more general class of quasilinear elliptic equations. *Journal of Differential equations*, 51(1):126–150, 1984.
- [34] Gary M Lieberman. Boundary regularity for solutions of degenerate elliptic equations. *Nonlinear Analysis: Theory, Methods & Applications*, 12(11):1203–1219, 1988.
- [35] Michael G Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1):1–67, 1992.
- [36] Adam M Oberman. Finite difference methods for the infinity laplace and p-laplace equations. *Journal of Computational and Applied Mathematics*, 254:65–80, 2013.
- [37] Guy Barles and Panagiotis E Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic analysis*, 4(3):271–283, 1991.
- [38] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. Graph Laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8(Jun):1325–1368, 2007.
- [39] Matthias Hein. Uniform convergence of adaptive graph-based regularization. In *International Conference on Computational Learning Theory*, pages 50–64. Springer, 2006.
- [40] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. In *International Conference on Computational Learning Theory*, pages 486–500. Springer, 2005.
- [41] Evarist Giné, Vladimir Koltchinskii, et al. Empirical graph laplacian approximation of Laplace–Beltrami operators: Large sample results. In *High dimensional probability*, pages 238–259. Institute of Mathematical Statistics, 2006.
- [42] Matthias Hein, Jean-Yves Audibert, and Ulrike Von Luxburg. From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians. In *International Conference on Computational Learning Theory*, pages 470–485. Springer, 2005.

- [43] Amit Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006.
- [44] Daniel Ting, Ling Huang, and Michael I Jordan. An analysis of the convergence of graph Laplacians. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 1079–1086, 2010.
- [45] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
- [46] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.
- [47] Sergei Bernstein. On a modification of Chebyshev’s inequality and of the error formula of Laplace. *Ann. Sci. Inst. Sav. Ukraine, Sect. Math*, 1(4):38–49, 1924.
- [48] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.
- [49] Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.
- [50] Petri Juutinen, Peter Lindqvist, and Juan J Manfredi. On the equivalence of viscosity solutions and weak solutions for a quasi-linear equation. *SIAM Journal on Mathematical Analysis*, 33(3):699–717, 2001.
- [51] Jeff Calder, Selim Esedoğlu, and Alfred O Hero III. A PDE-based approach to non-dominated sorting. *SIAM Journal on Numerical Analysis*, 53(1):82–104, 2015.
- [52] Hitoshi Ishii and Gou Nakamura. A class of integral equations and approximation of p-Laplace equations. *Calculus of Variations and Partial Differential Equations*, 37(3-4):485–522, 2010.

A Consistency for the graph p-Laplacian

Consider the graph p -Laplacian

$$(66) \quad \Delta_p^G u(x) := \sum_{y \in \mathcal{X}_n} w_n(x, y) |u_n(x) - u_n(y)|^{p-2} (u_n(x) - u_n(y)).$$

It is possible to prove Theorem 1 for the graph p -Laplacian provided $nh_n^p \rightarrow 0$ as $n \rightarrow \infty$. The proof is very similar to Theorem 1; the main difference is the consistency result, which we sketch here for completeness.

We require an integration lemma from [52].

Lemma 5. *Let $\rho : [0, \infty) \rightarrow [0, \infty)$ be a nonnegative continuous function with compact support, and let p_1, \dots, p_d be real numbers equal or larger than $1/2$. Then*

$$\int_{\mathbb{R}^d} \rho(|x|^2) |x_1|^{2p_1-1} \dots |x_d|^{2p_d-1} dx = \frac{\Gamma(p_1) \dots \Gamma(p_d)}{\Gamma(p_1 + \dots + p_d)} \int_0^\infty \rho(t) t^{p_1 + \dots + p_d - 1} dt,$$

where Γ denotes the Gamma function

$$\Gamma(t) = \int_0^\infty e^{-x} x^{t-1} dx.$$

We prove here consistency in expectation. Arguments similar to the proof of Theorem 5 can be used to obtain consistency with high probability and almost surely as $n \rightarrow \infty$.

Theorem 8. *Let $\varphi \in C^3(\mathbb{R}^d)$ and $p \geq 3$. Then*

$$(67) \quad \mathbb{E}[\Delta_p^G \varphi(x)] = \frac{1}{2} c(p, d) f^{-1} \operatorname{div}(f^2 |\nabla \varphi|^{p-2} \nabla \varphi) n h^{d+p} + R(x) n h^{d+p+1},$$

where

$$(68) \quad |R(x)| \leq C(c_1^{p-2}(c_1 + c_2 + c_3) + c_1^{p-3} c_2^3 h^2) \quad \text{for all } x \in \mathbb{R}^d,$$

and $c_k = \|D^k \varphi\|_{L^\infty(\mathbb{R}^d)}$.

Proof. We may assume $x = 0$ and $\varphi(0) = 0$. We write $h = h_n$ for simplicity. Then we have

$$\mathbb{E}[\Delta_p^G \varphi(0)] = n \int_{\mathbb{R}^d} \Phi_h(|y|)^p |\varphi(y)|^{p-2} \varphi(y) f(y) dy.$$

Set $x = y/h$ to find that

$$\mathbb{E}[\Delta_p^G \varphi(0)] = n h^d \int_{\mathbb{R}^d} \Phi(|x|)^p \Psi(\varphi(xh)) f(xh) dx,$$

where

$$\Psi(t) = |t|^{p-2} t.$$

Then

$$(69) \quad \Psi(t) = \Psi(a) + \Psi'(a)(t - a) + O(T^{p-3}|t - a|^3) \quad \text{for } a, t \in [-T, T].$$

Letting $q = \nabla \varphi(0)$ and $Q = \nabla^2 \varphi(0)$ we have

$$(70) \quad \varphi(xh) = O(c_1 |x| h),$$

$$(71) \quad \varphi(xh) = h q \cdot x + O(c_2 |xh|^2),$$

and

$$(72) \quad \varphi(xh) = hq \cdot x + \frac{h^2}{2}x \cdot Qx + O(c_3|xh|^3),$$

Therefore

$$\begin{aligned} \Psi(\varphi(xh)) &= \Psi(hq \cdot x) + \Psi'(hq \cdot x)(\varphi(xh) - hq \cdot x) + O(c_1^{p-3}|xh|^{p-3}c_2^3|xh|^6) \\ &= \Psi(hq \cdot x) + \Psi'(hq \cdot x) \left(\frac{h^2}{2}x \cdot Qx + O(c_3|x|^3h^3) \right) + O(c_1^{p-3}c_2^3|xh|^{p+3}) \\ &= \Psi(hq \cdot x) + \frac{h^2}{2}\Psi'(hq \cdot x)x \cdot Qx + O(c_1^{p-2}c_3|xh|^{p+1} + c_1^{p-3}c_2^3|xh|^{p+3}), \end{aligned}$$

and we have

$$\begin{aligned} \Psi(\varphi(xh))f(xh) &= f(0)\Psi(hq \cdot x) + f(0)\frac{h^2}{2}\Psi'(hq \cdot x)(x \cdot Qx) + (f(xh) - f(0))\Psi(hq \cdot x) \\ &\quad + O\left(c_1^{p-2}(c_2 + c_3)|xh|^{p+1} + c_1^{p-2}c_3|xh|^{p+2} + c_1^{p-3}c_2^3|xh|^{p+3} + c_1^{p-3}c_2^3|xh|^{p+4}\right) \\ &= f(0)\Psi(hq \cdot x) + f(0)\frac{h^2}{2}\Psi'(hq \cdot x)(x \cdot Qx) + h(\nabla f(0) \cdot x)\Psi(hq \cdot x) \\ &\quad + O\left(c_1^{p-2}(c_1 + c_2 + c_3)|xh|^{p+1} + c_1^{p-2}c_3|xh|^{p+2} + c_1^{p-3}c_2^3|xh|^{p+3} + c_1^{p-3}c_2^3|xh|^{p+4}\right). \end{aligned}$$

Integrating against $\Phi(|x|)^p$, the first term is odd and vanishes, so we get

$$\begin{aligned} &\int_{\mathbb{R}^d} \Phi(|x|)^p |\varphi(xh)|^{p-2} \varphi(xh) f(xh) dx \\ &= f(0)h^p \int_{\mathbb{R}^d} \Phi(|x|)^p \left(\left(\frac{1}{2}x \cdot Qx\right) \Psi'(q \cdot x) + (\nabla \log f(0) \cdot x) \Psi(q \cdot x) \right) dx \\ &\quad + O\left(c_1^{p-2}(c_1 + c_2 + c_3)h^{p+1} + c_1^{p-3}c_2^3h^{p+3}\right). \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\mathbb{E}[\Delta_p^G \varphi(0)]}{nh^{d+p}f(0)} &= \int_{\mathbb{R}^d} \Phi(|x|)^p \left(\left(\frac{1}{2}x \cdot Qx\right) \Psi'(q \cdot x) + (\nabla \log f(0) \cdot x) \Psi(q \cdot x) \right) dx \\ &\quad + O\left(c_1^{p-2}(c_1 + c_2 + c_3)h + c_1^{p-3}c_2^3h^3\right) \\ &= \int_{\mathbb{R}^d} \Phi(|x|)^p |q \cdot x|^{p-2} \left(\frac{1}{2}(p-1)(x \cdot Qx) + (\nabla \log f(0) \cdot x)(q \cdot x) \right) dx \\ &\quad + O\left(c_1^{p-2}(c_1 + c_2 + c_3)h + c_1^{p-3}c_2^3h^3\right) \\ (73) \quad &= A + B + O\left(c_1^{p-2}(c_1 + c_2 + c_3)h + c_1^{p-3}c_2^3h^3\right), \end{aligned}$$

where

$$(74) \quad A = \frac{p-1}{2} \sum_{i,j=1}^d Q_{ij} \int_{\mathbb{R}^d} \Phi(|x|)^p |q \cdot x|^{p-2} x_i x_j dx,$$

and

$$(75) \quad B = \sum_{i=1}^d (\log f)_{x_i} \int_{\mathbb{R}^d} \Phi(|x|)^p |q \cdot x|^{p-2} (q \cdot x) x_i dx.$$

Let us tackle A first. We may assume that $q \neq 0$. Let O be an orthogonal transformation so that $Oe_d = q/|q|$. In particular $O_{id} = [Oe_d]_i = q_i/|q|$. Making the change of variables $y = O^T x$ we have

$$\begin{aligned} B &= \sum_{i=1}^d (\log f)_{x_i} \int_{\mathbb{R}^d} \Phi(|y|)^p |q| |y_d|^{p-2} (|q| y_d) [Oy]_i dy \\ &= |q|^{p-1} \sum_{i,j=1}^d (\log f)_{x_i} O_{ij} \int_{\mathbb{R}^d} \Phi(|y|)^p |y_d|^{p-2} y_d y_j dy. \end{aligned}$$

If $j \neq d$ in the sum above, then the integral vanishes, since the integrand is odd. Therefore

$$(76) \quad B = |q|^{p-1} \sum_{i=1}^d (\log f)_{x_i} O_{id} \int_{B_2} \Phi(|y|)^p |y_d|^p dy = c(p, d) |q|^{p-2} (\nabla \log f) \cdot q,$$

where

$$c(p, d) = \int_{\mathbb{R}^d} \Phi(|x|)^p |x_d|^p dx.$$

For A , we again make the change of variables $y = O^T x$. Then we have

$$\begin{aligned} A &= \frac{p-1}{2} |q|^{p-2} \sum_{i,j=1}^d Q_{ij} \sum_{k,\ell=1}^d O_{ik} O_{j\ell} \int_{B_2} \Phi(|y|)^p |y_d|^{p-2} y_k y_\ell dy \\ &= \frac{p-1}{2} |q|^{p-2} \sum_{k=1}^d [O^T Q O]_{kk} \int_{B_2} \Phi(|y|)^p |y_d|^{p-2} y_k^2 dy \\ &= \frac{p-1}{2} |q|^{p-2} (\sigma(p, d) \text{Tr}(Q) + (c(p, d) - \sigma(p, d)) |q|^{-2} q \cdot Q q), \end{aligned}$$

where

$$\sigma(p, d) = \int_{\mathbb{R}^d} \Phi(|x|)^p |x_d|^{p-2} x_k^2 dx$$

for any $k < d$. By Lemma 5 we have

$$\sigma(p, d) = \frac{\Gamma\left(\frac{1}{2}\right)^{d-2} \Gamma\left(\frac{p-1}{2}\right) \Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{d+p}{2}\right)} \int_0^\infty \Phi(\sqrt{t}) t^{\frac{d+p}{2}-1} dt,$$

and

$$c(p, d) = \frac{\Gamma\left(\frac{1}{2}\right)^{d-1} \Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{d+p}{2}\right)} \int_0^\infty \Phi(\sqrt{t}) t^{\frac{d+p}{2}-1} dt.$$

Using the identity $\Gamma(t+1) = t\Gamma(t)$ we find that

$$c(p, d) = (p-1)\sigma(p, d).$$

It follows that

$$A = \frac{1}{2}c(p, d)|q|^{p-2}(\text{Tr}(Q) + (p-2)|q|^{-2}q^T Q q^T).$$

Combining this with (73) and (76) yields

$$\begin{aligned} & \frac{\mathbb{E}[\Delta_p^G \varphi(0)]}{nh^{d+p}} \\ &= \frac{1}{2}c(p, d)f|\nabla \varphi|^{p-2}(\Delta \varphi + 2\nabla \log f \cdot \nabla \varphi + (p-2)\Delta_\infty \varphi) \\ & \quad + O\left(c_1^{p-2}(c_1 + c_2 + c_3)h + c_1^{p-3}c_2^3h^3\right) \\ &= \frac{1}{2}c(p, d)f^{-1}\text{div}(f^2|\nabla \varphi|^{p-2}\nabla \varphi) + O\left(c_1^{p-2}(c_1 + c_2 + c_3)h + c_1^{p-3}c_2^3h^3\right). \end{aligned}$$

If $q = 0$ then from (73) we have

$$\frac{\mathbb{E}[\Delta_p^G \varphi(0)]}{nh^{d+p}} = O\left(c_1^{p-2}(c_1 + c_2 + c_3)h + c_1^{p-3}c_2^3h^3\right).$$

This completes the proof. □