Analysis of Semi-Supervised Learning on Hypergraphs

Adrien Weihs*[1], Andrea L. Bertozzi[1], and Matthew Thorpe[2]

[1]Department of Mathematics,
University of California Los Angeles,
Los Angeles, CA 90095, USA.

[2]Department of Statistics,
University of Warwick,
Coventry, CV4 7AL, UK.

November 2025

Abstract

Hypergraphs provide a natural framework for modeling higher-order interactions, yet their theoretical underpinnings in semi-supervised learning remain limited. We provide an asymptotic consistency analysis of variational learning on random geometric hypergraphs, precisely characterizing the conditions ensuring the well-posedness of hypergraph learning as well as showing convergence to a weighted p-Laplacian equation. Motivated by this, we propose Higher-Order Hypergraph Learning (HOHL), which regularizes via powers of Laplacians from skeleton graphs for multiscale smoothness. HOHL converges to a higher-order Sobolev seminorm. Empirically, it performs strongly on standard baselines.

Keywords and phrases. hypergraphs, non-parametric regression, semi-supervised learning, asymptotic consistency, multiscale problems
Mathematics Subject Classification. 49J55, 49J45, 62G20, 65N12

*Corresponding author. Email: weihs@math.ucla.edu

1 Introduction

Given a set of n feature vectors $\Omega_n = \{x_i\}_{i=1}^n \subset \Omega \subset \mathbb{R}^d$, where we assume $x_i \overset{\text{iid}}{\sim} \mu \in \mathcal{P}(\Omega)$, and a set of binary labels $\{y_i\}_{i=1}^N \subset \{0,1\}$, the goal of semi-supervised learning is to infer the missing labels for $\{x_i\}_{i=N+1}^n$ by leveraging both labeled and unlabeled data. Graph-based learning (see, e.g., [17,65] for non-deep learning methods and [55,95] for deep learning approaches) has become a widely used framework for this task, owing to the ability of graphs to effectively encode the underlying geometry of the data. Typically, one constructs a graph $G_n = (\Omega_n, W_n)$, where the nodes correspond to the feature vectors Ω_n and $W_n = (w_{ij})_{i,j=1}^n$ is an edge-weight matrix reflecting pairwise similarities. A variational problem is then posed on this graph to recover a labeling function on the vertices [8, 84, 94, 103]. When W_n is not intrinsic to the the data, the main subtlety lies in the edge-construction. For most applications, the assumption of graph homophily is made, i.e. connected points should share similar labels and therefore, one uses a local approach for the design of W_n. Most prominently, k-nearest neighbour graphs [18, 89] or random geometric graphs [26, 72, 84, 94] are employed.

We now detail the latter model. Given a parameter ε and a non-increasing function $\eta : [0,\infty) \mapsto [0,\infty)$, weights $w_{\varepsilon,ij}$ between vertices x_i and x_j are defined as

$$w_{\varepsilon,ij} = \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right).$$

The canonical choice of η is $\mathbb{1}_{[0,1]}$ which ensures that only points within an ε-distance are connected. In order to capture the geometry in a more delicate manner, one can go beyond pairwise interactions and decide to *more strongly connect* all points within the same ε-ball: e.g. if x_0, x_1, x_2 are all connected pairwise, then we could add an additional connection for the tuple (x_0, x_1, x_2). This observation leads to the considerations of hypergraphs (see Figure 1). Another alternative is, for example, to consider multiscale Laplace learning as in [65] and we refer to Section 3.3 for a thorough discussion about the link between multiscale Laplace learning and hypergraph learning.

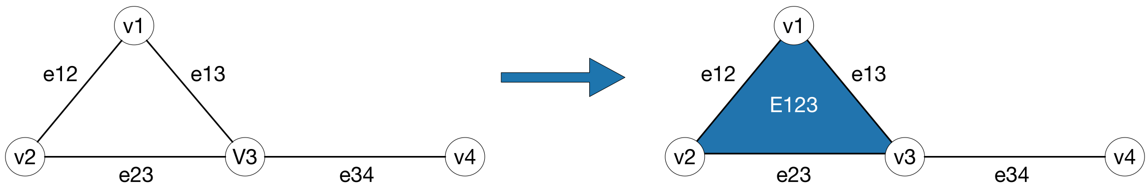

Figure 1: From graphs to hypergraphs. Left: In the graph, the vertices v_1, v_2, and v_3 are all connected pairwise. Right: A single hyperedge is added connecting all three vertices, transitioning from a graph to a hypergraph representation.

A hypergraph G is defined as $G = (V, E)$ where V is a set of objects and E a family of subsets e—called hyperedges—of V. Hypergraphs extend graphs by allowing hyperedges to connect arbitrary subsets of nodes, and hypergraph-based methods have found success in various areas of science as in [22, 30, 33, 34, 52, 62, 69, 73, 76, 78, 96, 98, 100]. We denote the weight of hyperedge e by $w_0(e)$ and its degree/size by $|e|$. In our setting, we will be considering $V = \Omega_n$. Learning on hypergraphs is, for example, considered in [100] where the following optimization problem (we keep the scaling with the edge degree but remove the vertex degree below) is proposed as a relaxation of the hypergraph cut problem (similar to what is done in [89] for the graph setting): feature vectors are clustered based on the value of the function u which solves

$$\text{(1)} \quad \operatorname*{argmin}_{v:\Omega_n\to\mathbb{R}} \sum_{e\in E} \sum_{\{x_i,x_j\}\subseteq e} \frac{w_0(e)}{|e|}(v(x_i) - v(x_j))^2 \quad \text{such that} \sum_{x_r\in\Omega_n} v(x_r)^2 = 1 \text{ and } \sum_{x_r\in\Omega_n} v(x_r) = 0.$$

Continuing the graph analogy, it is also shown that u is an eigenvector of the hypergraph Laplacian matrix. Similarly to what is considered in [103] (Laplace learning) or [84] (p-Laplacian learning), it is natural to extend (1) to the semi-supervised learning regime, namely to consider the solution to

$$\text{(2)} \quad \operatorname*{argmin}_{v:\Omega_n\to\mathbb{R}} \sum_{e\in E} \sum_{\{x_i,x_j\}\subseteq e} \frac{w_0(e,x_i,x_j)}{|e|}(v(x_i) - v(x_j))^2 \quad \text{such that } v(x_i) = y_i \text{ for } i \le N$$

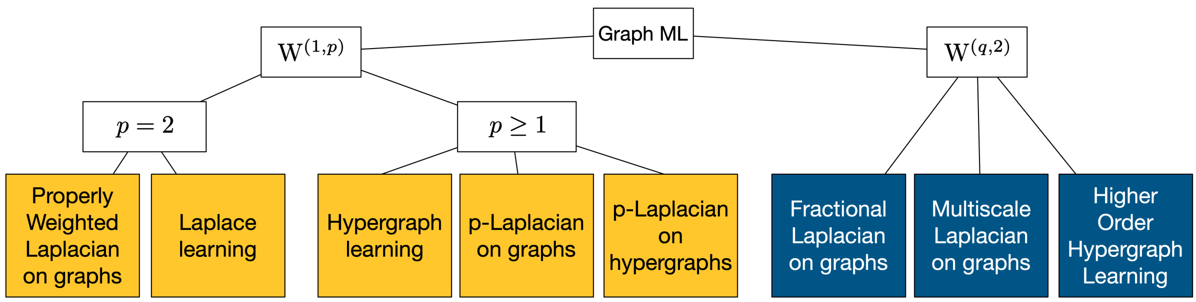

Figure 2: Classification of several algorithms based on their continuum limit. Edges indicate convergence to a Sobolev-type seminorm in the continuum limit, with each algorithm linked to its associated $\mathrm{W}^{(k,p)}$ space. Methods from left-to-right: [19], [103], [100], [27], [78], [26], [65], this work.

where the weight function w_0 now can also depend on the vertices x_i and x_j.

Several points are apparent from the formulation (2), which naturally decomposes into two components: the *mechanism* of label interaction, encoded via the finite difference term, and the *support* of these interactions, governed by the hypergraph's hyperedge weights and connectivity pattern.

P.1 Since the dependence on v in the objective function arises solely from pairwise interactions between vertices, the resulting regularization is inherently first-order. Consequently, the hypergraph itself contributes only to defining the support of interactions—that is, which and how strongly vertex pairs are connected—not to specifying the order or nature of the regularization being applied. We show in Theorems 3.2 and 3.3 that the solution of (2), in an appropriate limit, behaves as the sum of first order energies.

P.2 Given the previous point, the hyperedge construction is essential to the success of semi-supervised learning as in (2). We refer to [34] for a review of commonly used weight models in hypergraph learning.

P.3 In order to go beyond simple pairwise interaction, in Section 3.3, we propose another hypergraph learning model that penalizes higher-order derivatives similar to what can be found in [26, 94] for the graph setting. By doing so, we also establish a link between hypergraph learning and multiscale Laplace learning [65].

In order to make these insights rigorous, we provide a new consistency analysis of hypergraph learning (2) in the large-data regime. Asymptotic consistency analysis is a widely used framework for studying graph-based learning algorithms, see [7, 15, 24, 47, 82, 86]. It serves two main purposes: (1) identifying the continuum limit as $n \to \infty$ enables a principled classification of algorithmic behavior—see Figure 2 for a visual summary of several algorithms and their associated continuum limits; (2) analyzing the limiting continuum problem provides insights into fundamental properties of the discrete model, such as well-posedness and ill-posedness, i.e. convergence to nontrivial labelling functions or constants, on large datasets (see for example [26, 84, 94]). In contrast, consistency analysis for hypergraph-based regularization methods remains largely unexplored, with the exception of [78], and this work addresses that gap.

Consistency analysis between discrete energies $\mathcal{E}_n$ defined for discrete functions $v_n : \Omega_n \mapsto \mathbb{R}$ and an appropriate continuum energy $\mathcal{E}_\infty$ defined for continuum functions $v : \Omega \mapsto \mathbb{R}$, can be performed in various ways:

- in *spectral* convergence [7, 18, 35, 71, 83, 90, 91], one analyses the convergence of the eigenpairs of the discrete operator appearing in $\mathcal{E}_n$ to the eigenpairs of the corresponding continuum operator in $\mathcal{E}_\infty$;
- in *pointwise* convergence [7, 24, 47, 49, 51, 82, 86], one studies whether $\mathcal{E}_n(v\lfloor_{\Omega_n})$ converge to $\mathcal{E}_\infty(v)$ as $n \to \infty$ and $\varepsilon \to 0$, for a sufficiently smooth function $v : \Omega \to \mathbb{R}$—an alternative approach is to analyze the corresponding operators in the Euler–Lagrange equations and study their convergence instead [93];
- in *variational* convergence [15, 25, 26, 39–42, 84, 85, 87], one is interested in the convergence of minimizers of $\mathcal{E}_n$ to the minimizers of $\mathcal{E}_\infty$.

In this paper, we study the consistency of (2) through the lens of the latter two complementary notions of convergence. Pointwise convergence allows us to show that the operator appearing in the discrete Euler–Lagrange equation of (2) converges to a weighted p-Laplacian operator. Variational convergence, on the other hand, enables a precise characterization of the well- and ill-posedness of hypergraph learning in the semi-supervised setting, as a function of the hypergraph construction parameters. The appropriate analytical framework for convergence of minimizers is Γ-convergence [11], which requires a common metric space in which both discrete and continuum functions can be embedded. To this end, we rely on the TL^p space introduced in [39], which is based on tools from optimal transport [77, 88]. We review these methodologies in Section 2.

Another conclusion of our analysis is that hypergraph learning, in its standard variational formulation, is inherently a first-order method. This confirms our intuition discussed in Point **P.1**. Indeed, both the non-trivial operator obtained via pointwise convergence and the limiting energy identified through variational convergence are associated with first-order regularization: the former behaves like a p-Laplace operator, while the latter corresponds to a weighted Sobolev $\mathrm{W}^{1,p}$-seminorm. To move beyond first-order regularization and capture richer geometric structures encoded by the hypergraph, we therefore propose a new approach called Higher-Order Hypergraph Learning (HOHL) that penalizes higher-order derivatives of the labeling function. This is achieved by introducing powers of graph Laplacians computed at multiple scales, each induced by the underlying hypergraph structure.

We also study the variational convergence of the proposed HOHL model and provide a rigorous characterization of its well- and ill-posedness. This analysis also shows that HOHL converges to a genuinely higher-order Sobolev-type energy in the continuum limit, thereby distinguishing it fundamentally from classical hypergraph learning frameworks (see Figure 2) and addressing Point **P.3**. Furthermore, we establish a formal connection between HOHL and multiscale Laplacian learning, showing that our method can be interpreted as a principled extension of existing multiscale approaches. We support this interpretation with empirical results that demonstrate the effectiveness of HOHL in leveraging higher-order regularization for semi-supervised learning.

1.1 Contributions

Our main contributions (summarized in Table 1) are as follows:

1. **Pointwise Convergence to Continuum Operator:** We establish the quantitative pointwise convergence of the discrete Euler–Lagrange operators arising in hypergraph learning to a weighted p-Laplacian operator.

2. **Well-/Ill-Posedness via Variational Convergence:** We characterize the well- and ill-posedness of classical hypergraph learning via variational convergence, explicitly as a function of the hypergraph construction parameters.

3. **Classical Hypergraph Learning is First-Order:** We demonstrate that both the limiting operator and the limiting energy in classical hypergraph learning correspond to first-order regularization, thereby confirming that this framework is inherently first-order and effectively reduces to reweighted graph-based learning.

4. **HOHL and Multiscale Higher-Order Regularization:** We introduce the HOHL framework, which explicitly penalizes higher-order derivatives of the labeling function. Furthermore, we establish a formal link between HOHL and multiscale Laplacian learning and characterize the well-/ill-posedness of this model via variational convergence.

5. **Empirical Validation of HOHL:** We empirically validate HOHL on standard semi-supervised learning benchmarks, demonstrating its ability to exploit hypergraph geometric information.

6. **Continuum-Based Framework for Algorithm Classification:** We propose a principled framework for comparing and classifying graph- and hypergraph-based learning algorithms via their continuum limits. This asymptotic perspective reveals structural similarities between methods that may appear distinct in the discrete setting, as illustrated in Figure 2.

	Classical Hypergraph learning (2)	HOHL (5)
Pointwise convergence	Theorem 3.2	—
Well/Ill-posedness characterization	Theorem 3.3	Theorem 3.4
First-order method	✓	—
Higher-order method	—	✓

Table 1: Summary of contributions.

1.2 Related works

Laplace learning [103] is a widely used algorithm defined on undirected graphs and has served as a foundational tool for label propagation and semi-supervised learning. Its success has led to numerous extensions, including to directed graphs [50,99] and hypergraphs [100]. However, it has been observed that Laplace learning performs poorly in low-label regimes, especially when the number of unlabeled data points is large [27,68]. The asymptotic consistency analysis in [84] provides a theoretical explanation for this phenomenon and, by precisely characterizing the well-posedness of the algorithm, allows practitioners to circumvent limitations of Laplace learning (or of its variants – in the case of [84], the authors consider the more general p-Laplacian method). The work presented in this paper is partly concerned with the analogous analysis for the hypergraph learning algorithm of [100].

Our first contribution is to establish a continuum limit and characterize the well- and ill-posedness of hypergraph learning in the semi-supervised setting. This can be viewed as a continuation of recent efforts in discrete-to-continuum analysis of graph-based methods. For instance, the authors in [39] deal with the asymptotic consistency of graph total variation for clustering problems. Other examples include graph cut problems and their continuum counterparts in [37, 41, 42] and [70], the Mumford-Shah functional in [21] and an application in empirical risk minimization [36]. In the semi-supervised setting, the most relevant papers are [84, 94] dealing with p-Laplacian learning and fractional Laplacian regularization respectively. The latter two methods were developed as an alternative to Laplace learning [103], in order to loosen the strict well-posedness requirements of Laplace learning (other variants for this purpose are Lipschitz learning [14, 16, 58, 75], game theoretic p-Laplacian regularization [15], Poisson learning [13,17], re-weighting of the graph Laplacian in [19, 79–81] or truncated energies in [5, 6]). Extending these graph-based techniques to the hypergraph setting remains an important avenue for future research.

Our second contribution is an asymptotic characterization of the role of hypergraph structure in semi-supervised learning. Previous works demonstrated, that hypergraphs can offer meaningful advantages over graphs — see, for example, [69,97], where the richer combinatorial structure of hypergraphs is argued to better capture complex geometric relationships. This has led to the development of a variety of hypergraph-based learning methodologies [30, 34, 52, 62, 76, 100], often accompanied by discrete probabilistic or combinatorial comparisons with graph-based analogues (e.g., [3, 23, 53, 54, 67]). In contrast, our work focuses on the asymptotic regime, where $n \to \infty$, and studies the continuum limits of hypergraph learning objectives. To the best of our knowledge, such asymptotic analysis has only been carried out in [78], and only for their specific hypergraph regularization model. Extending this line of analysis to other hypergraph learning methodologies mentioned above remains an important and largely open direction. We believe that such asymptotic characterizations are highly valuable: as illustrated in Figure 2, identifying continuum limits provides a principled framework for comparing and classifying a broad range of algorithms that may appear quite different at the discrete level. Notably, our analysis reveals that certain hypergraph models, while distinct at the discrete level, asymptotically behave like graph-based models, revealing underlying structural equivalences.

We note that asymptotic analysis of hypergraphs has also been studied in the context of stochastic block models, where the hypergraph is itself a random object sampled from a generative process conditioned on the node labels [45, 46]. This setting is fundamentally different from ours which focuses on the regularization of label functions on a fixed hypergraph.

Our proposed HOHL model builds on prior work on nonlocal Laplacian-based regularization [26, 65, 94, 102], and introduces higher-order penalties by applying powers of the graph Laplacian at multiple scales induced by hypergraph structure. This unifies and generalizes ideas that have emerged in the graph neural network literature [2, 44, 66]. Although (hyper)graph neural networks (see [9] for a recent review) achieve strong empirical performance, they typically require extensive supervision, hyperparameter tuning, and often lack theoretical guarantees. In contrast, HOHL provides a theoretically grounded, architecture-free alternative that is well-suited for low-label regimes.

On the technical side, our proofs rely on nonlocal approximation results for $\mathrm{W}^{1,p}$ initially formulated in [10] and then considered under the Γ-convergence lens in [74]. The latter were extended to the discrete-to-continuum setting in [39, 84]. For other discrete-to-continuum results approximating $\mathrm{W}^{s,2}$ we refer to [26, 38, 94]. Finally, we note that our variational convergence results are purely asymptotic. Discrete-to-continuum rates have for example been obtained on similar graph learning problems in [20, 28, 93].

The remainder of the paper is structured as follows: in Section 2, we review the theoretical foundations of our approach; in Section 3, we state our main results; in Section 4, we provide detailed proofs; in Section 5, we present numerical experiments; in Section 6, we conclude and propose directions for future research.

2 Background

In this section, we introduce the TL^p topology and provide a brief overview of Γ-convergence. These two notions are essential for studying variational convergence, which concerns the convergence of minimizers of our machine learning objective. This perspective is particularly relevant in the semi-supervised learning setting, where the minimizers are often the primary objects of interest. In practice, label predictions are frequently obtained by thresholding the minimizer: for instance, if the discrete minimizer u_n takes real values and the observed labels satisfy $\{y_i\}_{i=1}^N \subset \{0,1\}$, then for $i > N$, predicted labels are set via thresholding as $y_i = \mathbb{1}_{[0,0.5]}(u_n(x_i))$. For more than two classes the methodology is extended using 1-hot encoding.

2.1 The TL^p Space

Let $\mathcal{P}(\Omega)$ be the set of probability measures on Ω and $\mathcal{P}_p(\Omega)$ be the set of probability measures on Ω with finite pth-moment. We denote by $\mathrm{L}^p(\mu)$ the set of functions u that are measurable with respect to μ and such that $\int_\Omega |u(x)|^p \,\mathrm{d}x < +\infty$. The pushforward of a measure $\mu \in \mathcal{P}(\Omega)$ by a map $T : \Omega \to \mathcal{Z}$ is the measure $\nu \in \mathcal{P}(\mathcal{Z})$ defined by

$$\nu(A) = T_{\#}\mu(A) := \mu(T^{-1}(A)) = \mu\left(\{x \,|\, T(x) \in A\}\right) \qquad \text{for all measurable sets } A.$$

For $\mu, \nu \in \mathcal{P}_p(\Omega)$ we denote by $\Pi(\mu,\nu)$ the set of all probability measures on $\Omega \times \Omega$ such that the first marginal is μ and the second marginal is ν, i.e. $(P_X)_{\#}\pi = \mu$ and $(P_Y)_{\#}\pi = \nu$ where $P_X : \Omega \times \Omega \ni (x,y) \mapsto x \in \Omega$ and $P_Y : \Omega \times \Omega \ni (x,y) \mapsto y \in \Omega$. The following definition of the TL^p space and metric can be found in [39].

Definition 2.1. *For an underlying domain Ω, define the set*

$$\mathrm{TL}^p = \{(\mu, u) \,|\, \mu \in \mathcal{P}_p(\Omega), u \in \mathrm{L}^p(\mu)\}\,.$$

For $(\mu,u),(\nu,v) \in \mathrm{TL}^p$, we define the TL^p distance d_{TL^p} as follows:

$$d_{\mathrm{TL}^p}((\mu,u),(\nu,v)) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{\Omega\times\Omega} |x-y|^p + |u(x) - v(y)|^p \,\mathrm{d}\pi(x,y) \right)^{\frac{1}{p}}.$$

The TL^p distance is related to the p-Wasserstein [77, 88] distance between the measures μ and ν and we refer to [39] for more details. In particular, from the latter property, we can characterize convergence in the TL^p space as follows.

Proposition 2.2. *[39, Proposition 3.12] Let $(\mu,u) \in \mathrm{TL}^p$ where μ is absolutely continuous with respect to Lebesgue measure and let $\{(\mu_n,u_n)\}_{n=1}^\infty$ be a sequence in TL^p. The following are equivalent:*

1. *(μ_n,u_n) converges to (μ,u) in TL^p;*

2. *μ_n converges weakly to μ and there exists a sequence of transport maps $\{T_n\}_{n=1}^{\infty}$ with $(T_n)_{\#}\mu = \mu_n$ and $\int_\Omega |x - T_n(x)|\,\mathrm{d}x \to 0$ such that*

$$\int_\Omega |u(x) - u(T_n(x))|^p \,\mathrm{d}\mu(x) \to 0;$$

In order to compare discrete functions to continuum ones, we will let μ_n be the empirical measures of our samples $\{x_i\}_{i=1}^n$ and μ will be the measure from which the points are sampled. Furthermore, $\{u_n\}_{n=1}^{\infty}$ and u will respectively be the minimizers of our discrete and continuum objectives. In order to use the above result, we need certain transport maps T_n whose existence is guaranteed by the next theorem [43, Theorem 1.1].

Theorem 2.3 (Existence of transport maps)**.** *Let $\Omega \subset \mathbb{R}^d$ be open, connected and bounded with Lipschitz boundary. Let μ be a probability measure on Ω with a density that is bounded above and below by positive constants. Let $x_i \overset{\text{iid}}{\sim} \mu \in \mathcal{P}(\Omega)$ and we denote the empirical measure of $\{x_i\}_{i=1}^n$ by μ_n. Then, there exists a constant $C > 0$ such that $\mathbb{P}$-a.s., there exists a sequence of transport maps $\{T_n : \Omega \mapsto \Omega_n\}_{n=1}^{\infty}$ from μ to μ_n such that:*

$$\begin{cases} \limsup_{n\to\infty} \frac{n^{1/2}\|\mathrm{Id}-T_n\|_{\mathrm{L}^\infty}}{\log(\log(n))} \leq C & \text{if } d = 1; \\ \limsup_{n\to\infty} \frac{n^{1/2}\|\mathrm{Id}-T_n\|_{\mathrm{L}^\infty}}{\log(n)^{3/4}} \leq C & \text{if } d = 2; \\ \limsup_{n\to\infty} \frac{n^{1/d}\|\mathrm{Id}-T_n\|_{\mathrm{L}^\infty}}{\log(n)^{1/d}} \leq C & \text{if } d \geq 3. \end{cases}$$

The probability measure $\mathbb{P}$ is defined in Section 3. In terms of the assumptions we introduce later, the conditions in the above theorem are given by **S.1**, **M.1**, **M.2** and **D.1**.

2.2 Γ-Convergence

The appropriate framework to describe the convergence of variational problems is Γ-convergence from the calculus of variations. We only recall the key properties used in this paper similarly to what be found in [39, 84, 94] and refer to [11] for more details.

Definition 2.4. *Let (Z, d_Z) be a metric space and $F_n : Z \to \mathbb{R}$ a sequence of functionals. We say that F_n Γ-converges to F with respect to d_Z if:*

1. *For every $z \in Z$ and every sequence $\{z_n\}$ with $d_Z(z_n, z) \to 0$:*
$$\liminf_{n\to\infty} F_n(z_n) \geq F(z);$$
2. *For every $z \in Z$, there exists a sequence $\{z_n\}$ with $d_Z(z_n, z) \to 0$ and*
$$\limsup_{n\to\infty} F_n(z_n) \leq F(z).$$

The notion of Γ-convergence allows one to derive the convergence of minimizers from compactness.

Definition 2.5. *We say that a sequence of functionals $F_n : Z \to \mathbb{R}$ has the compactness property if the following holds: if $\{n_k\}_{k\in\mathbb{N}}$ is an increasing sequence of integers and $\{z_k\}_{k\in\mathbb{N}}$ is a bounded sequence in Z for which $\sup_{k\in\mathbb{N}} F_{n_k}(z_k) < \infty$, then the closure of $\{z_k\}$ has a convergent subsequence.*

Proposition 2.6. *Convergence of minimizers. Let $F_n : Z \mapsto [0, \infty]$ be a sequence of functionals which are not identically equal to ∞. Suppose that the functionals satisfy the compactness property and that they Γ-converge to $F : Z \mapsto [0, \infty]$. Then*

$$\lim_{n\to\infty} \inf_{z\in Z} F_n(z) = \min_{z\in Z} F(z).$$

Furthermore, the closure of every bounded sequence $\{z_n\}$ for which

$$\lim_{n\to\infty} \left(F_n(z_n) - \inf_{z\in Z} F_n(z) \right) = 0 \tag{3}$$

has a convergent subsequence and each of its cluster points is a minimizer of F. In particular, if F has a unique minimizer, then any sequence satisfying (3) *converges to the unique minimizer of F.*

In this paper, we show that our discrete objectives Γ-converge (with respect to the TL^p-topology) to the appropriate continuum objectives. Then, we will show that the sequence of discrete minimizers are precompact in TL^p and, using Proposition 2.6, deduce that the latter converge to the continuum minimizers.

3 Main results

In this section, we present our main results as well the relevant notation and assumptions used for our proofs.

3.1 General notation

For $z \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d\times d}$, we denote by $(z)_i$ the i-th coordinate of z and by $(A)_{ij}$ the ij-th element of A. We denote the Sobolev space of functions in L^p with k-th order derivatives as $\mathrm{W}^{k,p}$ [61].

3.2 Hypergraph setting

We will use the same probabilistic setting as in [94]. In particular, the idea is to consider a probability space with measure $\mathbb{P}$ in which elements are sequences $\{x_i\}_{i=1}^\infty$. Our results will be formulated in terms of $\mathbb{P}$, showing that certain properties holds for a set Ψ of sequences $\{x_i\}_{i=1}^\infty$ with $\mathbb{P}(\Psi) = 1$.

Given a set of n feature vectors $\Omega_n = \{x_i\}_{i=1}^n \subset \Omega \subset \mathbb{R}^d$ where we assume that $x_i \overset{\text{iid}}{\sim} \mu \in \mathcal{P}(\Omega)$, a length-scale $\varepsilon > 0$ and a function $\eta : [0,\infty) \mapsto [0,\infty)$, we can define weights $w_{\varepsilon,ij}$ between vertices x_i and x_j as follows:

$$w_{\varepsilon,ij} = \eta\left(\frac{|x_i - x_j|}{\varepsilon}\right).$$

The graph $(\Omega_n, W_{n,\varepsilon})$ where $W_{\varepsilon,n} = \{w_{\varepsilon,ij}\}_{i,j=1}^n$ is called a random geometric graph [72].

We now define essential matrices related to such graphs. Let $D_{n,\varepsilon}$ be the diagonal matrix with entries $d_{n,\varepsilon,ii} = \sum_{j=1}^n w_{\varepsilon,ij}$ and define $\sigma_\eta = \frac{1}{d}\int_{\mathbb{R}^d} \eta(|h|)|h|^2 \,\mathrm{d}h < \infty$. The graph Laplacian is defined as

$$\Delta_{n,\varepsilon} := \frac{2}{\sigma_\eta n \varepsilon^{d+2}}(D_{n,\varepsilon} - W_{n,\varepsilon}).$$

The latter can be interpreted as a matrix $\Delta_{n,\varepsilon} \in \mathbb{R}^{n\times n}$ or as an operator $\Delta_{n,\varepsilon} : \mathrm{L}^2(\mu_n) \to \mathrm{L}^2(\mu_n)$ where $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ is the empirical measure.

Given functions $u_n, v_n : \Omega_n \to \mathbb{R}$, we also define the $\mathrm{L}^2(\mu_n)$ inner product:

$$\langle u_n, v_n\rangle_{\mathrm{L}^2(\mu_n)} = \frac{1}{n}\sum_{i=1}^n u_n(x_i)v_n(x_i).$$

Such functions can be considered vectors in $\mathbb{R}^n$ and we will understand u_n as both a function $u_n : \Omega_n \to \mathbb{R}$ and a vector $\mathbb{R}^n$.

The eigenpairs of $\Delta_{n,\varepsilon}$ are denoted by $\{(\lambda_{n,\varepsilon,k}, \psi_{n,\varepsilon,k})\}_{k=1}^n$ where $\lambda_{n,\varepsilon,k}$ are in increasing order, $0 = \lambda_{n,\varepsilon,1} < \lambda_{n,\varepsilon,2} \le \lambda_{n,\varepsilon,3} \le \ldots \le \lambda_{n,\varepsilon,n}$, (where strict inequality between $\lambda_{n,\varepsilon,1}$ and $\lambda_{n,\varepsilon,2}$ follows when the graph $G_{n,\varepsilon}$ is connected) and we note that $\{\psi_{n,\varepsilon,k}\}_{k=1}^n$ form for a basis of $\mathrm{L}^2(\mu_n)$.

We can generalize the random geometric graph weight model to create random geometric hypergraphs. In particular, we define the weight of a hyperedge of size $k+1$ by aggregating pairwise interactions as

$$w_{\varepsilon,i_0\cdots i_k} = \prod_{j=1}^k \prod_{r=0}^{j-1} w_{\varepsilon,i_j i_r}. \tag{4}$$

This construction biases the model toward hyperedges whose constituent nodes lie within a shared neighborhood, effectively encoding a finer notion of locality. For instance, choosing $\eta = \mathbf{1}_{[0,1]}$ yields $w_{\varepsilon,i_0\cdots i_k} > 0$ if and only if the entire tuple $(x_{i_0}, \ldots, x_{i_k})$ lies within a common ball of radius ε. In this sense, ε should be thought of as the length-scale of interaction between vertices. We denote by $t(k)$ the number of terms in the product $\prod_{j=1}^k \prod_{r=0}^{j-1} w_{\varepsilon,i_j i_r}$.

The weight construction introduced in (4) serves as the foundation for all subsequent theoretical analysis. In particular, we reformulate the hypergraph learning energy (2) using this weight model in (7). With $\eta_{\mathrm{p}}(x_{i_0}, \ldots, x_{i_k}) = \prod_{j=1}^k \prod_{r=0}^{j-1} \eta\left(\frac{|x_{i_j} - x_{i_r}|}{\varepsilon}\right)$, we define the discrete (k,p)-Laplacian operators which we relate to the hypergraph learning energy (7) in Proposition 3.1 as

$$\Delta_{n,\varepsilon}^{(k,p)}(u)(x_{i_0}) = \frac{1}{n^k \varepsilon^{p+kd}} \sum_{i_1,\ldots,i_k=1}^n \left[\eta_{\mathrm{p}}(x_{i_0}, \ldots, x_{i_k})|u(x_{i_1}) - u(x_{i_0})|^{p-2}(u(x_{i_1}) - u(x_{i_0}))\right].$$

We note that the $(1,2)$-Laplacian is just $\Delta_{n,\varepsilon}$.

In order to introduce the continuum counterpart of $\Delta_{n,\varepsilon}^{(k,p)}$, we first define

$$\tilde{\eta}_{\mathrm{p}}(z_1,\dots,z_k) = \left[\prod_{s=1}^{k} \eta(|z_s|)\right] \left[\prod_{j=2}^{k}\prod_{r=1}^{j-1} \eta(|z_j - z_r|)\right],$$

and the constant

$$\sigma_\eta^{(k,p)} = \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)|(\tilde{z}_1)_d|^p \,\mathrm{d}\tilde{z}_k \cdots \mathrm{d}\tilde{z}_1.$$

In Theorem 3.2, we will establish the precise asymptotic relationship between $\Delta_{n,\varepsilon}^{(k,p)}$ and the following operator:

$$\Delta_\infty^{(k,p)}(u)(x) = \frac{\sigma_\eta^{(k,p)}}{2\rho(x)} \mathrm{div}(\rho(x)^{k+1}\|\nabla u\|_2^{p-2}\nabla u(x)).$$

3.3 Higher-order hypergraph learning

To penalize higher-order derivatives in a discrete setting, one must go beyond pairwise interactions in the energy functional (as in (2)). In this section, we introduce the Higher-Order Hypergraph Learning (HOHL) model, which is specifically designed to achieve this. In particular, we recall that $v^\top \Delta_{n,\varepsilon}^s v$, with $s \in \mathbb{R}$, corresponds to a discrete Sobolev $\mathrm{W}^{s,2}$ semi-norm applied to v [26, 94].

Let (V,E) be a hypergraph (independently of its weight model), and define $q = \max_{e\in E}|e| - 1$ as the maximum hyperedge size minus one. For each $k \in \{1,\dots,q\}$, we define a corresponding skeleton graph $G^{(k)} = (V, E^{(k)})$ by

$$E^{(k)} = \{\{v_i, v_j\} \mid \exists e \in E \text{ with } |e| = k+1 \text{ and } \{v_i,v_j\} \subset e\},$$

that is, $G^{(k)}$ includes all pairwise edges induced by hyperedges of size $k+1$. Let $L^{(k)}$ denote the graph Laplacian associated with $G^{(k)}$. In hypergraph models where larger hyperedges connect increasingly closer points—a locality principle satisfied by our random model defined via (4)—the skeleton graphs $G^{(k)}$ become more selective as k increases, capturing finer local interactions.

We define the HOHL energy as

$$v^\top \left[\sum_{k=1}^{q} \lambda_k (L^{(k)})^{p_k}\right] v \tag{5}$$

for $v \in \mathbb{R}^n$, where $0 < p_1 < \ldots < p_q$ are powers and $\lambda_1,\dots,\lambda_q > 0$ are tuning parameters. In practice, we often set $p_k = k$ for simplicity, although the same reasoning applies to any positive and increasing sequence $\{p_k\}_{k=1}^q$. For hypergraph models based on the locality principle detailed above, this energy imposes a hierarchical, scale-aware regularization: small k enforces global smoothness, while large k imposes fine-grained regularity on hyperedges of large sizes. Figure 3 illustrates this mechanism. In this paper, we restrict our attention to the HOHL energy (5) applied to the random hypergraph model where the vertex set is $\Omega_n \subset \mathbb{R}^d$ and hyperedges are constructed via (4). For a generalization of (5) to non-geometric datasets and arbitrary hypergraphs, as well as an analysis of the computational properties of HOHL, we refer the reader to [92].

The idea behind HOHL—imposing higher-order regularity based on sample closeness—naturally reduces, in point cloud settings where locality is geometrically defined, to a multiscale approach. In particular, our framework provides a theoretical justification for the multiscale Laplace model proposed in [65]. In the latter, the authors consider the energy

$$v^\top \left[\sum_{k=1}^{q} \lambda_k \Delta_{n,\varepsilon^{(k)}}^{p_k}\right] v, \tag{6}$$

where $\varepsilon^{(1)} > \cdots > \varepsilon^{(q)}$, and $p_k > 0$ controls the regularity imposed at each scale. When using the random hypergraph model, our HOHL formulation recovers this multiscale behavior via hypergraph structure and offers

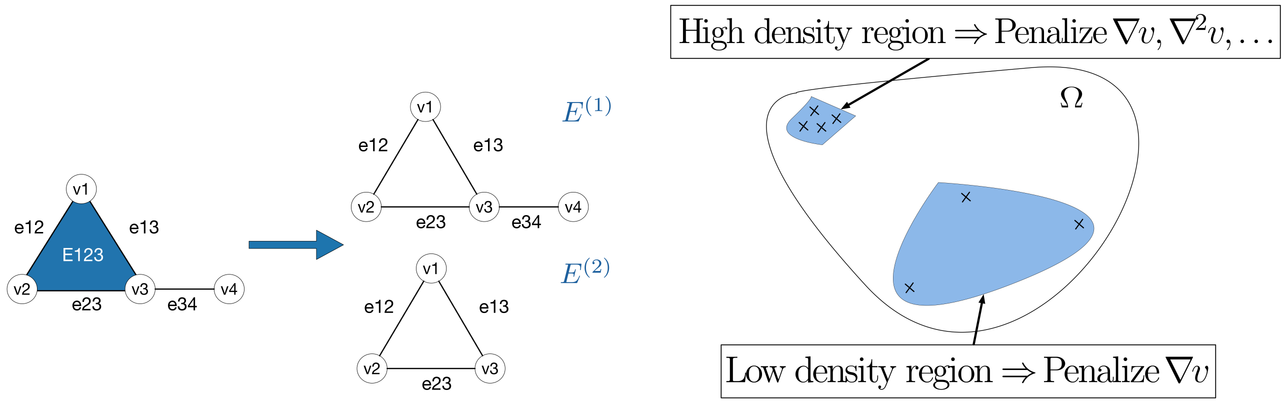

Figure 3: Illustration of the HOHL energy with $p_k = k$. Left: For $q = 2$, the energy imposes hierarchical regularization by penalizing $v^\top L^{(1)} v$ on skeleton edges $E^{(1)}$ and $v^\top (L^{(2)})^2 v$ on $E^{(2)}$. Right: With the random hypergraph model of (4), in high-density regions, hyperedges of large size capture finer structural details, and HOHL imposes stronger smoothness to exploit this local structure.

a principled explanation for choosing $p_k = k$ (or any increasing sequence): since $v^\top \Delta^s_{n,\varepsilon} v$ converges (under suitable conditions) to a $\mathrm{W}^{s,2}$ seminorm in the continuum, increasing p_k enforces higher regularity in high-density regions where large hyperedges—and hence skeleton graphs with large k—are more likely to form. The main distinction between (5) and (6) lies in how the Laplacians are obtained: in (6), they are constructed explicitly from multiple scaled kernels, while in (5), they are derived from the hypergraph structure.

While the hypergraph construction in (4) is theoretically appealing, it becomes computationally expensive due to the combinatorial growth in hyperedge enumeration. To address this, we propose using the multiscale Laplace model (6), which preserves the hierarchical regularization structure of HOHL while remaining computationally efficient and analytically tractable. Although (6) does not exactly approximate (5) with weights from (4)—since the corresponding limiting Laplacians may differ in density scaling—it provides a practical surrogate for point clouds embedded in a metric space, where hypergraphs are constructed based on proximity and we do not have access to the (skeleton) Laplacians. In contrast, for general hypergraphs where the Laplacians $L^{(k)}$ are known, we apply (5) directly.

3.4 Variational problems for hypergraph learning

For some $p > 1$ and $k \geq 1$, the classical hypergraph energy can be written as

(7) $$\mathcal{E}^{(k,p)}_{n,\varepsilon}(v) = \frac{1}{n^{k+1}\varepsilon^{p+kd}} \sum_{i_0,\cdots,i_k=1}^{n} \left[\prod_{j=1}^{k} \prod_{r=0}^{j-1} \eta\left(\frac{|x_{i_j} - x_{i_r}|}{\varepsilon} \right) \right] |v(x_{i_1}) - v(x_{i_0})|^p$$

for $v : \Omega \to \mathbb{R}$ while the associated discrete semi-supervised learning objective is

$$\mathcal{F}^{(k,p)}_{n,\varepsilon}((\nu, v)) = \begin{cases} \mathcal{E}^{(k,p)}_{n,\varepsilon}(v) & \text{if } \nu = \mu_n \text{ and for } i \leq N, v(x_i) = y_i \\ +\infty & \text{else} \end{cases}$$

for $(\nu, v) \in \mathrm{TL}^p(\Omega)$ and where $\{y_i\}_{i=1}^N \subset \{0, 1\}$ are binary labels.

For ρ the density of μ with respect to Lebesgue measure, in the continuum, we define

(8) $$\begin{aligned} \mathcal{E}^{(k,p)}_{\infty}(v) &= \int_\Omega \int_{(\mathbb{R}^d)^k} \left[\prod_{s=1}^{k} \eta(|z_s|) \right] \left[\prod_{j=1}^{k} \prod_{r=1}^{j-1} \eta(|z_j - z_r|) \right] |\nabla v(x_0) \cdot z_1|^p \rho(x_0)^{k+1} \, \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0 \\ &= \int_{(\mathbb{R}^d)^k} \left[\prod_{s=1}^{k} \eta(|z_s|) \right] \left[\prod_{j=1}^{k} \prod_{r=1}^{j-1} \eta(|z_j - z_r|) \right] |e \cdot z_1|^p \, \mathrm{d}z_k \cdots \mathrm{d}z_1 \int_\Omega \|\nabla v(x_0)\|_2^p \rho(x_0)^{k+1} \, \mathrm{d}x_0 \\ &=: \sigma^{(k)}_\eta \int_\Omega \|\nabla v(x_0)\|_2^p \rho(x_0)^{k+1} \, \mathrm{d}x_0 \end{aligned}$$

where $e \in \mathbb{R}^d$ is any vector with $\|e\|_2 = 1$ and (8) follows by isotropy of the kernels. The corresponding semi-supervised learning objectives are:

$$\mathcal{F}_\infty^{(k,p)}((\nu, v)) = \begin{cases} \mathcal{E}_\infty^{(k,p)}(v) & \text{if } \nu = \mu,\ v \in \mathrm{W}^{1,p}(\Omega) \text{ and for } i \leq N, v(x_i) = y_i \\ +\infty & \text{else,} \end{cases}$$

$$\mathcal{G}_\infty^{(k,p)}((\nu, v)) = \begin{cases} \mathcal{E}_\infty^{(k,p)}(v) & \text{if } \nu = \mu \text{ and } v \in \mathrm{W}^{1,p}(\Omega) \\ +\infty & \text{else.} \end{cases}$$

Our final objective, for $q \geq 1$ and a positive sequence $\{\lambda_k\}_{k=1}^q \subseteq \mathbb{R}$, is to consider the sums

$$(\mathcal{SF})_{n,\varepsilon}^{(q,p)}((\nu, v)) = \sum_{k=1}^q \lambda_k \mathcal{F}_{n,\varepsilon}^{(k,p)}((\nu, v)),$$

$$(\mathcal{SF})_\infty^{(q,p)}((\nu, v)) = \sum_{k=1}^q \lambda_k \mathcal{F}_\infty^{(k,p)}((\nu, v))$$

and

$$(\mathcal{SG})_\infty^{(q,p)}((\nu, v)) = \sum_{k=1}^q \lambda_k \mathcal{G}_\infty^{(k,p)}((\nu, v)).$$

3.5 Variational problems for higher-order hypergraph learning

For HOHL and $p > 0$, we define the discrete energies

$$\mathcal{I}_{n,\Delta_{n,\varepsilon}}^{(p)}(v) = \langle v, \Delta_{n,\varepsilon}^p v \rangle_{\mathrm{L}^2(\mu_n)}$$

for $v : \Omega \to \mathbb{R}$ and their associated semi-supervised learning objectives

$$\mathcal{J}_{n,\Delta_{n,\varepsilon}}^{(p)}((\nu, v)) = \begin{cases} \mathcal{I}_{n,\Delta_{n,\varepsilon}}^{(p)}(v) & \text{if } \nu = \mu_n \text{ and for } i \leq N, v(x_i) = y_i \\ +\infty & \text{else} \end{cases}$$

for $(\nu, v) \in \mathrm{TL}^p(\Omega)$.

The latter have continuum analogues. Indeed, let Δ_ρ be the continuum weighted Laplacian operator defined by

$$\Delta_\rho u(x) = -\frac{1}{\rho(x)} \mathrm{div}(\rho^2 \nabla u)(x),\ x \in \Omega \qquad \frac{\partial u}{\partial n} = 0,\ x \in \partial\Omega$$

and let $\{(\beta_i, \psi_i)\}_{i=1}^\infty$ be its associated eigenpairs where $\beta_1 = 0 < \beta_2 \leq \beta_3 \leq \ldots$. We note that $\{\psi_i\}_{i=1}^\infty$ form a basis of $\mathrm{L}^2(\mu)$. The continuum energy is then defined as

$$\mathcal{I}_\infty^{(p)}(v) = \langle v, \Delta_\rho^p v \rangle_{\mathrm{L}^2(\mu)}$$

for $v : \Omega \to \mathbb{R}$ and we consider the following well-posed and ill-posed semi-supervised learning objectives:

$$\mathcal{J}_\infty^{(p)}((\nu, v)) = \begin{cases} \mathcal{I}_\infty^{(p)}(v) & \text{if } \nu = \mu,\ v \in \mathcal{H}^p(\Omega) \text{ and for } i \leq N, v(x_i) = y_i \\ +\infty & \text{else,} \end{cases}$$

$$\mathcal{K}_\infty^{(p)}((\nu, v)) = \begin{cases} \mathcal{I}_\infty^{(p)}(v) & \text{if } \nu = \mu \text{ and } v \in \mathcal{H}^p(\Omega) \\ +\infty & \text{else} \end{cases}$$

for $(\nu, v) \in \mathrm{TL}^p(\Omega)$ and where

$$\mathcal{H}^p(\Omega) = \{h \in \mathrm{L}^2(\mu) \,|\, \mathcal{I}_\infty^{(p)}(h) < +\infty\}.$$

The latter set can be shown to be very closely related to the Sobolev space $W^{p,2}(\Omega)$ [26, Lemma 17]. Finally, for $q \geq 1$ and positive sequences $\{\lambda_k\}_{k=1}^q \subseteq \mathbb{R}$, $P := \{p_k\}_{k=1}^q \subseteq \mathbb{R}$ and $E := \{\varepsilon^{(k)}\}_{k=1}^q$ with $\varepsilon^{(1)} > \cdots > \varepsilon^{(q)}$, we consider the sums

$$(\mathcal{SJ})_{n,E}^{(q,P)}((\nu,v)) = \sum_{k=1}^{q} \lambda_k \mathcal{J}_{n,\Delta_{n,\varepsilon^{(k)}}}^{(p_k)}((\nu,v)),$$

$$(\mathcal{SJ})_{\infty}^{(q,P)}((\nu,v)) = \sum_{k=1}^{q} \lambda_k \mathcal{J}_{\infty}^{(p_k)}((\nu,v))$$

and

$$(\mathcal{SK})_{\infty}^{(q,P)}((\nu,v)) = \sum_{k=1}^{q} \lambda_k \mathcal{K}_{\infty}^{(p_k)}((\nu,v)).$$

We will also index our length-scales by the number of vertices, i.e. $\varepsilon^{(k)} = \varepsilon_n^{(k)}$, and in this case, we write $E_n := \{\varepsilon_n^{(k)}\}_{k=1}^q$. The above sums correspond to the multiscale model for HOHL as detailed in Section 3.3.

3.6 Assumptions

In this section, we list the assumptions used throughout the paper.

Assumptions 1. Assumption on the space. We assume either **S.1** or **S.2**.

S.1 The feature vector space Ω is an open, connected and bounded subset of $\mathbb{R}^d$ with Lipschitz boundary.

S.2 The feature vector space Ω is the unit torus $\mathbb{R}^d/\mathbb{Z}^d$.

Assumptions 2. Assumptions on the measure. In most cases we need both **M.1** and **M.2**.

M.1 The measure μ is a probability measure on Ω.

M.2 There is a continuous Lebesgue density ρ of μ which is bounded from above and below by strictly positive constants, i.e. $0 < \min_{x\in\Omega} \rho(x) \leq \max_{x\in\Omega} \rho(x) < +\infty$.

The data consists of feature vectors $\{x_i\}_{i=1}^n$ and labels $\{y_i\}_{i=1}^N$ and we make the following assumptions.

Assumptions 3. Assumptions on the data. Assumption **D.1** is needed for consistency results and **D.2** is needed in the semi-supervised setting.

D.1 Feature vectors $\Omega_n = \{x_i\}_{i=1}^n$ are iid samples from a measure μ satisfying **M.1**.

D.2 There are N labels $\{y_i\}_{i=1}^N \subset \mathbb{R}$ corresponding to the first N feature vectors $\{x_i\}_{i=1}^N$.

The weight function η is assumed to satisfy the following assumptions.

Assumptions 4. Assumptions on the weight function or kernel.

W.1 The function $\eta : [0,\infty) \to [0,\infty)$ is non-increasing, has compact support, is continuous and positive at $x = 0$.

The compactness of the support of η corresponds to the setting in most applications where, for computational purposes, one wants to restrict the range of interactions between vertices in our hypergraph. Theoretically however, the compact support assumption is not strictly necessary and we can extend our results to the non-compactly supported case as in done in [39, 84].

Finally, we make the following assumption on the length scale $\varepsilon = \varepsilon_n$ which we scale with n.

Assumptions 5. Assumptions on the length-scale. For our consistency results we will need one of **L.1**, **L.2** or **L.3**.

L.1 The length scale $\varepsilon = \varepsilon_n$ is positive, converges to 0, i.e. $0 < \varepsilon_n \to 0$.

L.2 The length scale $\varepsilon = \varepsilon_n$ is positive, converges to 0, i.e. $0 < \varepsilon_n \to 0$, and satisfies the following lower bound:

$$\lim_{n\to\infty} \frac{\log(n)}{n\varepsilon_n^d} = 0 \qquad \text{if } d \geq 3;$$
$$\lim_{n\to\infty} \frac{(\log(n))^{3/2}}{n\varepsilon_n^2} = 0 \qquad \text{if } d = 2;$$
$$\lim_{n\to\infty} \frac{(\log(\log(n))}{n\varepsilon_n^2} = 0 \qquad \text{if } d = 1.$$

L.3 The length scale $\varepsilon = \varepsilon_n$ is positive, converges to 0, i.e. $0 < \varepsilon_n \to 0$ and satisfies the following lower bound:

$$\lim_{n\to\infty} \frac{\log(n)}{n\varepsilon_n^{d+4}} = 0.$$

Assumption **L.2** guarantees that (with probability one – measured with $\mathbb{P}$) that there exists N_1 such that for all $n \geq N_1$ the graph $G_{n,\varepsilon_n} = (\Omega_n, W_{n,\varepsilon_n})$ is connected (see [48] or [72]). We also note that the condition in the $d = 2$ case in Assumption **L.2** can be tightened by removing the log-term (using the techniques from [18,21]), i.e. $\lim_{n\to\infty} \frac{\log(n)}{n\varepsilon_n^d} = 0$ for $d \geq 2$, so that ε_n can be chosen to be any sequence asymptotically greater than the connectivity radius for all $d \geq 2$.

3.7 Main results

We give our results for (classical) hypergraph learning (2) in Subsection 3.7.1, and our results for higher order hypergraph learning (6) in Subsection 3.7.2.

3.7.1 Hypergraph learning

We start by determining the corresponding Euler–Lagrange equations of (7). The result implies that the gradient of the energy decomposes into a sum of discrete operators, each tied to a hyperedges of a given size. The proof is in Section 4.1.

Proposition 3.1 (Discrete Euler-Lagrange equations of hypergraph learning)**.** *The energy*

$$v \mapsto \sum_{k=1}^{q} \lambda_k \mathcal{E}_{n,\varepsilon}^{(k,p)}(v)$$

is minimized by u if and only if u satisfies $\sum_{k=1}^q \lambda_k \Delta_{n,\varepsilon_n}^{(k,p)}(u) = 0$.

We now study the asymptotic behavior as $n \to \infty$. The next result shows quantitative pointwise convergence of the discrete hypergraph operator to its continuum analogue.

Theorem 3.2 (Pointwise consistency)**.** *Assume that Assumptions **S.1**, **M.1**, **M.2**, **D.1** and **L.1** hold. Furthermore, assume that $\rho \in \mathrm{C}^2(\Omega)$. Let Ω' be compactly contained in Ω, $q \geq 1$, $\{\lambda_k\}_{k=1}^q \subset (0,\infty)$, $p \in \{2\} \cup [3,\infty)$, $\varepsilon_n \leq \delta$ and $u \in \mathrm{C}^3$. Then, for n large enough, we have that*

$$\left| \left(\sum_{k=1}^{q} \lambda_k \Delta_{n,\varepsilon_n}^{(k,p)} \right)(u)(x_{i_0}) - \rho(x_{i_0}) \left(\sum_{k=1}^{q} \lambda_k \Delta_{\infty}^{(k,p)} \right)(u)(x_{i_0}) \right| = \mathcal{O}\left(\delta \|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1} \right)$$

for $x_{i_0} \in \Omega_n \cap \Omega'$, with probability $1 - Cn\exp\left(-Cn\varepsilon_n^{2(1+qd)}\delta^2\right)$ where $C > 0$ is a constant independent of n and δ.

The proof uses a Taylor expansion of $|t|^{p-2}t$ for $p \geq 3$, relying on bounded second derivatives. For $p = 2$, the expansion is exact. The details are in Section 4.2.

Next, we precisely characterize the asymptotic consistency of hypergraph learning as a function of the length-scale ε_n. We refer to Figure 4 for a visual summary of the result.

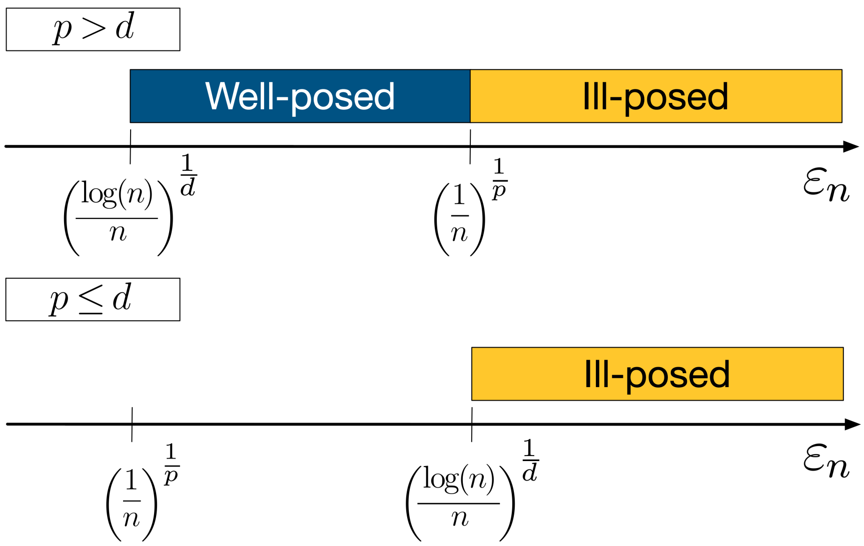

Figure 4: Well- and Ill-posedness characterization of hypergraph learning as a function of the length-scale ε_n.

Theorem 3.3 (Asymptotic consistency analysis of hypergraph learning)**.** *Assume that* ***S.1****,* ***M.1****,* ***M.2****,* ***D.1****,* ***D.2****,* ***W.1****, and* ***L.2*** *hold. Let (μ_n, u_n) be minimizers of $(\mathcal{SF})^{(q,p)}_{n,\varepsilon_n}$.*

1. *(Well-posed case) Assume that $n\varepsilon_n^p \to 0$. Then, $\mathbb{P}$-a.s., there exists a continuous function u such that $(\mu_n, u_n) \to (\mu, u)$ in $\mathrm{TL}^p(\Omega)$ and for any $\Omega' \subset\subset \Omega$, $\max_{\{r \le n \,|\, x_r \in \Omega'\}} |u(x_r) - u_n(x_r)| \to 0$. In particular, (μ, u) is a minimizer of $(\mathcal{SF})^{(q,p)}_{\infty}$.*
2. *(Ill-posed case) Assume that $n\varepsilon_n^p \to \infty$. Then, $\mathbb{P}$-a.s., there exists $u \in \mathrm{W}^{1,p}(\Omega)$ and a subsequence $\{n_r\}_{r=1}^{\infty}$ such that $(\mu_{n_r}, u_{n_r}) \to (\mu, u)$ in $\mathrm{TL}^p(\Omega)$ and (μ, u) is a minimizer of $(\mathcal{SG})^{(q,p)}_{\infty}$.*

The limiting energy identified in Theorem 3.3 is

$$\sum_{k=1}^{q} \lambda_k \sigma_\eta^{(k)} \int_\Omega \|\nabla v(x_0)\|_2^p \, \rho(x_0)^{k+1} \, \mathrm{d}x_0 = \int_\Omega \|\nabla v(x_0)\|_2^p \left(\sum_{k=1}^{q} \lambda_k \sigma_\eta^{(k)} \rho(x_0)^{k+1} \right) \mathrm{d}x_0. \tag{9}$$

In particular, it only differs from the limiting energy of p-Laplacian learning

$$\min_{v:\Omega \mapsto \mathbb{R}} \int \|\nabla v(x_0)\|_2^p \rho(x_0)^2 \, \mathrm{d}x_0$$

by a weighting of the density. This also means that, asymptotically, these hypergraph models are similar to graph models. We thus obtain an identical characterization of well-/ill-posedness in terms of ε_n as in [84, Theorem 2.1], now in the hypergraph context. In particular, the well-posedness, in which minimizers smoothly interpolate the known labels, is ensured if and only if ε_n satisfies the lower bound **L.2** and the upper bound $n\varepsilon_n^p \to 0$.

Similarly to [94, Remark 3.1], the above-mentioned bounds also imply that $p > d$ and we recover an intuition stemming from Sobolev spaces. Indeed, in the continuum, our functions in $\mathrm{W}^{1,p}(\Omega)$ must be at least continuous in order to satisfy pointwise constraints, i.e. be in the well-posed case: by Sobolev inequalities, this can only the case whenever $p > d$. Our results show that this condition is necessary but not sufficient as ε_n also has to satisfy an upper bound. We also note that in practice, the condition $p > d$ often leads to $p \geq 3$, which satisfies the requirements for pointwise convergence in Theorem 3.2.

For the ill-posed case, we note that minimizers of $(\mathcal{SG})^{(q,p)}_{\infty}$ are constants and therefore, for large n, we expect our discrete minimizers to be almost constant with spikes at the known labels (this is observed for $q = 1$ in [27, 68]). The labelling problem relying on the thresholding of our minimizers is therefore rendered nonsensical, hence our denomination of ill-posed. The case $p \leq d$ is also covered by our characterization of our ill-posed case (see [94, Remark 3.3]), linking our results back to the Sobolev embedding intuition.

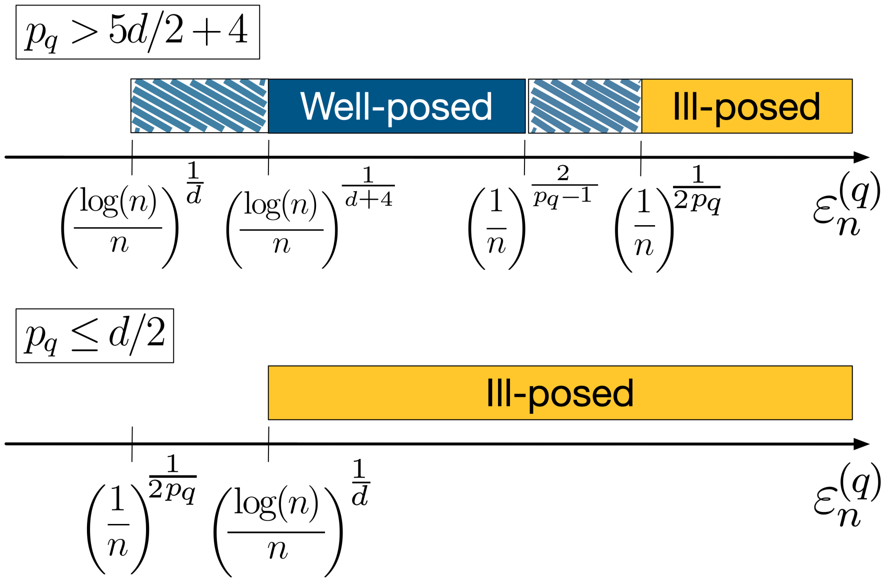

Figure 5: Well- and Ill-posedness characterization of HOHL as a function of the length-scale $\varepsilon_n^{(q)}$. The striped regions are conjectured results.

3.7.2 Higher Order Hypergraph Learning

We observe from the definition of $\Delta_\infty^{(k,p)}$ and the limiting energy functional (9) that the notion of locality encoded by hyperedges of size $k+1$ appears through higher powers of the sampling density ρ. This has the beneficial effect of amplifying the contribution of high-density regions during learning. As mentioned in the introduction, this illustrates that hypergraph learning—at least in the classical formulation—modifies where pairwise interactions occur, but not how they occur. In contrast, the HOHL framework fundamentally alters the nature of interactions by introducing higher-order terms. For instance, the term $v^\top (L^{(2)})^2 v$ incorporates nested finite differences of $L^{(2)}(x_i)$ (see [89]), which already represent aggregated information from multiple neighbors. Such terms approximate second-order derivatives (see Section 3.3) and cannot be decomposed into purely pairwise interactions. In this way, HOHL better leverages the hypergraph structure by simultaneously modifying the support of interactions (via multiscale decompositions) and the mechanism of interaction (through higher-order regularization).

We obtain the following asymptotic consistency result for HOHL/the multiscale Laplacian learning energy (6). Figure 5 provides a visual summary of the result.

Theorem 3.4 (Asymptotic consistency analysis of higher-order hypergraph learning)**.** *Assume that* ***S.2****,* ***M.1****,* ***M.2****,* ***D.1****,* ***D.2*** *and* ***W.1*** *hold. Let (μ_n, u_n) be minimizers of $(\mathcal{SJ})_{n,\varepsilon_n}^{(q,p)}$. Let $q \geq 1$, $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$ with $p_1 \leq \cdots \leq p_q$ and $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$ with $\varepsilon_n^{(1)} > \cdots > \varepsilon_n^{(q)}$. Assume that $\rho \in \mathrm{C}^\infty$.*

1. *(Well-posed case) Assume that $\varepsilon_n^{(q)}$ satisfies* ***L.3****, that $n \cdot (\varepsilon_n^{(q)})^{p_q/2-1/2}$ is bounded and that $p_q > \frac{5}{2}d + 4$. Then, $\mathbb{P}$-a.s., there exists a continuous function u such that $(\mu_n, u_n) \to (\mu, u)$ in $\mathrm{TL}^2(\Omega)$ and $\max_{\{r \leq n\}} |u(x_r) - u_n(x_r)| \to 0$. In particular, (μ, u) is a minimizer of $(\mathcal{SJ})_\infty^{(q,p)}$.*

2. *(Ill-posed case) Assume that $\varepsilon_n^{(q)}$ satisfies* ***L.2*** *as well as $n(\varepsilon_n^{(q)})^{2p_q} \to \infty$. Furthermore, assume that $\sup_{n\geq 1} \|u_n\|_{\mathrm{L}^2(\mu_n)}$ is bounded. Then, $\mathbb{P}$-a.s., there exists u and a subsequence $\{n_r\}_{r=1}^\infty$ such that $(\mu_{n_r}, u_{n_r}) \to (\mu, u)$ in $\mathrm{TL}^2(\Omega)$ and (μ, u) is a minimizer of $(\mathcal{SK})_\infty^{(q,p)}$.*

Similarly to Theorem 3.3, this result shows how the choice of scale and regularity governs the transition between expressive interpolation (i.e. minimizers are smooth-functions satisfying the pointwise constraints in the well-posed case) and trivial smoothing (i.e. minimizers are constants in the ill-posed case). Theorem 3.4 identifies the precise parameter regime needed for well-posedness in semi-supervised learning with HOHL. Notably, we remark that the characterization mostly depends on the parameters of the finest scale, i.e. $\varepsilon_n^{(q)}$ and p_q.

In contrast to standard hypergraph learning, which converges to a $\mathrm{W}^{1,p}$ seminorm, as explained in Section 3.3, the continuum limiting enery identified through Theorem 3.4 indicates that HOHL converges to a $\mathrm{W}^{p_q,2}$ seminorm, as illustrated in Figure 2. This underscores the distinct regularity structure induced by our higher-order formulation.

Furthermore, the same Sobolev intuition developed for hypergraph learning prevails for HOHL. In particular, our result implies that $p_q > d/2$—or equivalently that $\mathrm{W}^{p_q,2}$ is embedded in C^0—is necessary for well-posedness. Similarly $p_q \leq d/2$ also partly characterizes the ill-posed case.

4 Proofs

4.1 Euler-Lagrange equations of hypergraph learning

In this section, we present the proof for the derivation of the Euler-Lagrange equations of hypergraph learning.

Proof of Proposition 3.1. We want to obtain the Euler-Lagrange equation of $\sum_{k=1}^q \lambda_k \mathcal{E}_{n,\varepsilon}^{(k,p)}(\cdot)$ and to this purpose, let us evaluate

$$\left(\frac{d}{dt} \sum_{k=1}^q \lambda_k \mathcal{E}_{n,\varepsilon}^{(k,p)}(u+tv) \right) \lfloor_{t=0} = \sum_{k=1}^q \lambda_k \frac{d}{dt} \mathcal{E}_{n,\varepsilon}^{(k,p)}(u+tv) \lfloor_{t=0}.$$

In particular, we proceed as follows:

$$\begin{aligned}
T &:= \frac{d}{dt} \mathcal{E}_{n,\varepsilon}^{(k,p)}(u+tv)\lfloor_{t=0} \\
&= \frac{p}{n^{k+1}\varepsilon^{p+kd}} \sum_{i_0,\cdots,i_k=1}^n \eta_{\mathrm{p}}(x_{i_0},\dots,x_{i_k}) \left| u(x_{i_1}) - u(x_{i_0}) \right|^{p-2} (u(x_{i_1}) - u(x_{i_0}))(v(x_{i_1}) - v(x_{i_0})) \\
&=: \sum_{i_0,i_1=1}^n g(x_{i_0},x_{i_1})(v(x_{i_1}) - v(x_{i_0})) \\
&= \sum_{i_0,i_1=1}^n g(x_{i_0},x_{i_1})v(x_{i_1}) - \sum_{i_0,i_1=1}^n g(x_{i_0},x_{i_1})v(x_{i_0}) \\
&= \sum_{i_0,i_1=1}^n g(x_{i_1},x_{i_0})v(x_{i_0}) - \sum_{i_0,i_1=1}^n g(x_{i_0},x_{i_1})v(x_{i_0}) \\
&= -2 \sum_{i_0,i_1=1}^n g(x_{i_0},x_{i_1})v(x_{i_0}) \qquad (10) \\
&= \langle -2p\Delta_{n,\varepsilon}^{(k,p)}(u), v \rangle_{\mathrm{L}^2(\mu_n)}
\end{aligned}$$

where we used the fact that the function $f(x,y) = \eta_{\mathrm{p}}(x,y,x_{i_2},\dots,x_{i_k})$ satisfies $f(x,y) = f(y,x)$ for all fixed $x_{i_2},\dots,x_{i_k}$ implying that $g(x,y) = -g(y,x)$ for (10). From this, we deduce that u minimizing $\sum_{k=1}^q \lambda_k \mathcal{E}_{n,\varepsilon}^{(k,p)}$ must satisfy

$$\sum_{k=1}^q \lambda_k \Delta_{n,\varepsilon_n}^{(k,p)}(u) = 0.$$

Conversely, by convexity any u satisfying $\sum_{k=1}^q \lambda_k \Delta_{n,\varepsilon_n}^{(k,p)}(u) = 0$ must be a minimizer. □

4.2 Pointwise convergence of hypergraph learning

In this section, we present the proofs related to Theorem 3.2.

4.2.1 Preliminary results

We start with several auxiliary results. First, the following inequality [64] will be useful in the proof of Theorem 3.2.

Theorem 4.1 (McDiarmid/Azuma Inequality)**.** *Let $X_1, \dots, X_n$ be iid random variables satisfying $|X_i| \leq M$ almost surely. Let $Y_n = f(X_1, \dots, X_n)$ for some function f. If there exists $b > 0$ such that f satisfies*

$$|f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, \tilde{x}_i, \dots, x_n)| \leq b$$

for all x_i and $\tilde{x}_i$, $1 \leq i \leq n$, then for all $t > 0$,

$$\mathbb{P}(|Y_n - \mathbb{E}(Y_n)| \geq t) \leq 2 \exp\left(-\frac{t^2}{2nb^2}\right).$$

Lemma 4.2. *Let $S^{(n,k)}(i) = \#\{(\alpha_1, \dots, \alpha_k) \in \{1, \dots, n\}^k \,|\, \exists 1 \leq \ell \leq k$ such that $\alpha_\ell = i\}$. Then, for $1 \leq i \leq n$, $S^{(n,k)}(i) = n^{k-1} + (n-1)S^{(n,k-1)}(i)$ and $S^{(n,k)}(i) = \leq kn^{k-1}$.*

Proof. Let $(\alpha_1, \dots, \alpha_k) \in \{1, \dots, n\}^k$. If we fix $\alpha_1 = i$, then there exists n^{k-1} tuples of the form $(i, \dots, \alpha_k)$. Now, if $\alpha_1 = j \neq i$, there exist $S^{(n,k-1)}(i)$ tuples of the form $(j, \alpha_2, \dots, \alpha_k)$ that contain at least one i. Since, j can take $n-1$ values, we conclude that $S^{(n,k)}(i) = n^{k-1} + (n-1)S^{(n,k-1)}(i)$. The second claim can be proven simply by induction. □

The following lemma can be proven by induction and Taylor's expansion.

Lemma 4.3 (Product identity)**.** *Let $\rho \in \mathrm{C}^2(\mathbb{R}^d)$. Then, for $k \geq 1$, we have*

$$\prod_{\ell=1}^{k} \rho(x_{i_0} + \varepsilon_n z_\ell) = \rho(x_{i_0})^k + \varepsilon_n \rho(x_{i_0})^{k-1} \nabla\rho(x_{i_0})(z_1 + \cdots + z_k) + \mathcal{O}(\varepsilon_n^2)$$

for $x_{i_0}, z_1, \dots, z_k \in \mathbb{R}^d$ and $\varepsilon_n \in \mathbb{R}$.

We recall the following lemma from [93].

Lemma 4.4 (Asymptotics of domain of integration)**.** *Assume that $\Omega \subset \mathbb{R}^d$ is a bounded open domain. Let $\varepsilon_n > 0$ be a sequence that tends to 0, Ω' be compactly contained in Ω and $C \subset \mathbb{R}^d$ be a compact subset. Then, for n large enough, for all $x_{i_0} \in \Omega'$, the set $S_{\varepsilon_n}(x_{i_0}) = \{z \in \mathbb{R}^d \,|\, x_{i_0} + \varepsilon_n z \in \Omega\} \cap C$ is equal to C.*

We recall that $f : (\mathbb{R}^d)^k \mapsto \mathbb{R}$ is odd symmetric if $f(-x_1, \dots, -x_k) = -f(x_1, \dots, x_k)$ and that for such a function, $\int_A f(x_1, \dots, x_k)\, \mathrm{d}x_k \cdots \mathrm{d}x_1 = 0$ if A is symmetric.

4.2.2 Equivalent representation of the continuum Laplacian

In this section, we prove an equivalent representation of the continuum Laplacian $\Delta_\infty^{(k,p)}$. The latter will appear as the continuum limit in Theorem 3.2. We start by introducing the following constants:

$$\sigma_\eta^{(k,p,1)} = \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1, \dots, \tilde{z}_k) |(\tilde{z}_1)_d|^{p-2} (\tilde{z}_1)_1^2 \, \mathrm{d}\tilde{z}_k \cdots \mathrm{d}\tilde{z}_1,$$

and

$$\sigma_\eta^{(k,p,2)} = \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1, \dots, \tilde{z}_k) |(\tilde{z}_1)_d|^{p-2} (\tilde{z}_1)_d (\tilde{z}_2)_d \, \mathrm{d}\tilde{z}_k \cdots \mathrm{d}\tilde{z}_1.$$

The key idea for the following computations is to consider integrals of the form

$$\int_{(\mathbb{R}^d)^k} \widetilde{\eta}_p(z_1, \dots, z_k) g(z_1, \dots, z_k) \, \mathrm{d}z_1 \cdots \mathrm{d}z_k$$

as an expectation with respect to the measure $\mathbb{Q}$ defined through the density

$$f(z_1, \dots, z_k) = \frac{1}{\mathcal{Z}} \widetilde{\eta}_p(z_1, \dots, z_k), \qquad \mathcal{Z} := \int_{(\mathbb{R}^d)^k} \widetilde{\eta}_p(z) \, \mathrm{d}z.$$

By considering a random vector $(Z_1, \dots, Z_k) \sim \mathbb{Q}$, we obtain that

$$\mathcal{Z} \int_{(\mathbb{R}^d)^k} \frac{\widetilde{\eta}_p(z_1, \dots, z_k)}{\mathcal{Z}} g(z_1, \dots, z_k)\, dz_1 \cdots dz_k = \mathcal{Z}\, \mathbb{E}_{\mathbb{Q}}[g(Z_1, \dots, Z_k)].$$

The structure of $\widetilde{\eta}_p$ implies strong symmetry properties of the law of $(Z_1, \dots, Z_k)$, specifically invariance under simultaneous rotations of all coordinates and under affine reflections fixing Z_1. Exploiting these invariances via conditional expectations and multivariate symmetry arguments, we obtain the identities between $\sigma_\eta^{(k,p)}$, $\sigma_\eta^{(k,p,1)}$ and $\sigma_\eta^{(k,p,2)}$ stated below.

Lemma 4.5 (Radial marginal)**.** *Assume that Assumption* ***W.1*** *holds. Let $(Z_1, \dots, Z_k)$ be a random vector in $(\mathbb{R}^d)^k$ with distribution $\mathbb{Q}$ defined through the density*

$$f(z_1, \dots, z_k) = \frac{1}{\mathcal{Z}}\, \widetilde{\eta}_p(z_1, \dots, z_k), \qquad \mathcal{Z} := \int_{(\mathbb{R}^d)^k} \widetilde{\eta}_p(z)\, \mathrm{d}z.$$

Then, the marginal law of Z_1 is rotation-invariant.

Proof. Let $Q \in O(d)$ be any orthogonal matrix. Orthogonality implies that $\|Qx\| = \|x\|$ for all $x \in \mathbb{R}^d$, and more generally

$$\|Qx - Qy\| = \|x - y\|$$

for all $x, y \in \mathbb{R}^d$. Therefore each factor in $\widetilde{\eta}_p$ is invariant under the simultaneous rotation

$$(z_1, \dots, z_k) \mapsto (Qz_1, \dots, Qz_k).$$

Indeed,

$$\eta(\|Qz_s\|) = \eta(\|z_s\|), \qquad \eta(\|Qz_j - Qz_r\|) = \eta(\|z_j - z_r\|),$$

so

$$\widetilde{\eta}_p(Qz_1, \dots, Qz_k) = \widetilde{\eta}_p(z_1, \dots, z_k)$$

for all $z_1, \dots, z_k \in \mathbb{R}^d$. Since the Jacobian determinant of a rotation is 1, it follows that the probability density f satisfies

$$f(Qz_1, \dots, Qz_k) = f(z_1, \dots, z_k).$$

Thus the law of $(Z_1, \dots, Z_k)$ is rotation-invariant under simultaneous rotations of all coordinates:

$$(Z_1, \dots, Z_k) \overset{d}{=} (QZ_1, \dots, QZ_k).$$

By [12, Proposition 4.1.1], this implies that every marginal of $\mathbb{Q}$ is also rotation-invariant. □

Lemma 4.6 (Constant identity I)**.** *Assume that Assumption* (**W.1**) *holds. Let $d \geq 2$. Then,*

$$\sigma_\eta^{(k,p)} = (p-1)\, \sigma_\eta^{(k,p,1)}.$$

Proof. We define

$$\mathcal{Z} := \int_{(\mathbb{R}^d)^k} \widetilde{\eta}_p(z_1, \dots, z_k)\, \mathrm{d}z_1 \dots \mathrm{d}z_k$$

and

$$\mathrm{d}\mathbb{Q}(z_1, \cdots, z_k) := \frac{1}{\mathcal{Z}} \widetilde{\eta}_p(z_1, \dots, z_k)\, \mathrm{d}z_1 \dots \mathrm{d}z_k.$$

For $(Z_1, \dots, Z_k) \sim \mathbb{Q}$, we write $Z_1 = (X_1, \dots, X_d)^\top$. Then,

$$\sigma_\eta^{(k,p)} = \mathcal{Z}\, \mathbb{E}[|X_d|^p], \qquad \text{and} \qquad \sigma_\eta^{(k,p,1)} = \mathcal{Z}\, \mathbb{E}[|X_d|^{p-2} X_1^2]$$

where the expectation is taken with respect to $\mathbb{Q}$.

By Lemma 4.5, the marginal law of Z_1 is rotation-invariant. By [12, Theorem 4.1.2], since the distribution of Z_1 is rotation invariant, we may write

$$Z_1 \overset{d}{=} RU,$$

where $R \stackrel{d}{=} \|Z_1\| \geq 0$, $U \in \mathbb{S}^{d-1}$ is uniformly distributed on the unit sphere and R and U are independent. Writing $U = (U_1, \ldots, U_d)$, we have

$$|X_d|^p \stackrel{d}{=} R^p |U_d|^p \qquad \text{and} \qquad |X_d|^{p-2} X_1^2 \stackrel{d}{=} R^p |U_d|^{p-2} U_1^2.$$

Therefore, using the independence of R and U, we obtain that

$$\frac{\mathbb{E}[|X_d|^p]}{\mathbb{E}[|X_d|^{p-2}X_1^2]} = \frac{\mathbb{E}(R^p)\mathbb{E}[|U_d|^p]}{\mathbb{E}(R^p)\mathbb{E}[|U_d|^{p-2}U_1^2]} = \frac{\mathbb{E}[|U_d|^p]}{\mathbb{E}[|U_d|^{p-2}U_1^2]}.$$

Let $Y_i := U_i^2$. For U uniform on $\mathbb{S}^{d-1}$, by the proof of [29, Theorem 3.3] (which shows that $(U_1, \ldots, U_d) \stackrel{d}{=} x/\|x\|$ where $x \sim \mathcal{N}(0, \mathrm{Id}_{d\times d})$) and [29, Section 1.4] (which shows that $(x_1^2/\|x\|, \ldots, x_d^2/\|x\|)$ is Dirichlet-distributed with parameters $(\frac{1}{2}, \ldots, \frac{1}{2})$), the vector $(Y_1, \ldots, Y_d)$ is Dirichlet–distributed with parameters $(\frac{1}{2}, \ldots, \frac{1}{2})$. We can then apply the moment formula for Dirichlet distributions [57, Section 27.6]: for $Y \sim \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_n)$ and $\beta_i > 0$,

$$\mathbb{E}\left[\prod_{i=1}^n Y_i^{\beta_i}\right] = \frac{\Gamma\left(\sum_{i=1}^n \alpha_i\right)}{\Gamma\left(\sum_{i=1}^n \alpha_i + \beta_i\right)} \prod_{i=1}^n \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)}.$$

This yields:

$$\mathbb{E}[|U_d|^p] = \mathbb{E}[Y_d^{p/2}] = \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d+p}{2}\right)} \frac{\Gamma\left(\frac{1}{2}\right)^{d-1}\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^d} = \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma(\frac{d}{2})}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{d+p}{2}\right)}$$

and

$$\mathbb{E}[|U_d|^{p-2}U_1^2] = \mathbb{E}[Y_d^{(p-2)/2}Y_1] = \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d+p-2+2}{2}\right)} \frac{\Gamma\left(\frac{1}{2}\right)^{d-2}\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^d} = \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d+p}{2}\right)} \frac{\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^2}.$$

Taking the ratio, we obtain

$$\frac{\mathbb{E}[|U_d|^p]}{\mathbb{E}[|U_d|^{p-2}U_1^2]} = \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{3}{2}\right)} = \frac{\frac{p-1}{2}\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{p-1}{2}\right)\frac{1}{2}\Gamma\left(\frac{1}{2}\right)} = p - 1$$

where we used the identity $\Gamma(t+1) = t\Gamma(t)$ for the middle equality. We conclude that

$$\sigma_\eta^{(k,p)} = (p-1)\,\sigma_\eta^{(k,p,1)}.$$

□

Lemma 4.7 (Reflections)**.** *Let $z_1 \in \mathbb{R}^d$ be non-zero. Define $v := \frac{z_1}{\|z_1\|}$, $m := \frac{z_1}{2}$, the function $R_{z_1} : \mathbb{R}^d \to \mathbb{R}^d$ by*

$$R_{z_1}(y) := y - 2(y \cdot v)\, v,$$

and the function $S_{z_1} : \mathbb{R}^d \to \mathbb{R}^d$ by

$$S_{z_1}(y) := m + R_{z_1}(y - m).$$

Then,:

1. *R_{z_1} is the reflection across the hyperplane $\{y : y \cdot v = 0\}$ and an isometry;*
2. *S_{z_1} is the reflection across the affine hyperplane $H_{z_1} := \{y \in \mathbb{R}^d : (y - m) \cdot v = 0\}$;*
3. *S_{z_1} is an isometry;*
4. *$S_{z_1}(0) = z_1$ and $S_{z_1}(z_1) = 0$;*
5. *$\|S_{z_1}(y)\| = \|z_1 - y\|$ and $\|S_{z_1}(y) - z_1\| = \|y\|$.*

Proof. 1. R_{z_1} is linear by definition. We check that R_{z_1} is an isometry and has the expected geometric action. For any $y \in \mathbb{R}^d$, decompose

$$y = (y \cdot v)\, v + y_\perp,$$

where $y_\perp := y - (y \cdot v)v$ satisfies $y_\perp \cdot v = 0$. Then

$$\begin{aligned} R_{z_1}(y) &= y - 2(y \cdot v)\, v \\ &= (y \cdot v)v + y_\perp - 2(y \cdot v)v \\ &= -(y \cdot v)v + y_\perp. \end{aligned} \tag{11}$$

Thus, R_{z_1} flips the component along v and preserves the orthogonal component, which is precisely the reflection across the hyperplane $\{y : y \cdot v = 0\}$.

Moreover, by the orthogonality of $(y \cdot v)\, v$ and $y_\perp$, we have

$$\begin{aligned} \|R_{z_1}(y)\|^2 &= \| -(y \cdot v)v + y_\perp\|^2 \\ &= \|y_\perp\|^2 + (y \cdot v)^2 \\ &= \|(y \cdot v)v + y_\perp\|^2 \\ &= \|y\|^2, \end{aligned}$$

so that R_{z_1} is an isometry.

2. S_{z_1} fixes every point of the hyperplane H_{z_1}. Indeed, if $y \in H_{z_1}$, then $(y - m) \cdot v = 0$, and hence

$$R_{z_1}(y - m) = (y - m) - 2\big((y - m) \cdot v\big)v = y - m.$$

It follows that

$$S_{z_1}(y) = m + R_{z_1}(y - m) = m + (y - m) = y.$$

For a general point $y \in \mathbb{R}^d$, the vector $y - m$ has the orthogonal decomposition

$$y - m = \big((y - m) \cdot v\big)v + (y - m)_\perp, \qquad (y - m)_\perp \cdot v = 0.$$

Using the reflection identity (11), we obtain

$$R_{z_1}(y - m) = -(y - m) \cdot v\, v + (y - m)_\perp,$$

so R_{z_1} reverses the normal component $\big((y-m)\cdot v\big)v$ and preserves the tangential component $(y-m)_\perp$. Geometrically, R_{z_1} is the reflection across the hyperplane

$$H_0 := \{y \in \mathbb{R}^d : y \cdot v = 0\}$$

The hyperplane

$$H_{z_1} := \{y \in \mathbb{R}^d : (y - m) \cdot v = 0\}$$

is simply the translation of H_0 by the vector m. Therefore, to obtain the reflection across H_{z_1}, we must conjugate R by this translation, which yields the affine map

$$S_{z_1}(y) = m + R_{z_1}(y - m).$$

Thus S_{z_1} is precisely the affine reflection across H_{z_1} (see Figure 6).

3. For any $y, r \in \mathbb{R}^d$,

$$\begin{aligned} S_{z_1}(y) - S_{z_1}(r) &= \big(m + R_{z_1}(y - m)\big) - \big(m + R_{z_1}(r - m)\big) \\ &= R_{z_1}(y - m) - R_{z_1}(r - m) \\ &= R_{z_1}\big((y - m) - (r - m)\big) \\ &= R_{z_1}(y - r), \end{aligned} \tag{12}$$

where we used the linearity of R_{z_1} for (12). Since R_{z_1} is an isometry by part 1 of the lemma,

$$\|S_{z_1}(y) - S_{z_1}(r)\| = \|R_{z_1}(y - r)\| = \|y - r\|.$$

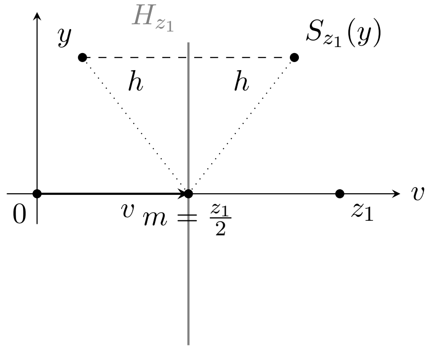

Figure 6: Geometric interpretation of the affine reflection S_{z_1} across the hyperplane H_{z_1} orthogonal to v and passing through $m = z_1/2$. The points y and $S_{z_1}(y)$ are symmetric with respect to H_{z_1}.

4. We have

$$\begin{aligned} R_{z_1}(-m) &= -m - 2\big((-m)\cdot v\big)v \\ &= -m - 2\Big(-\frac{z_1}{2}\cdot\frac{z_1}{\|z_1\|}\Big)\frac{z_1}{\|z_1\|} \\ &= -m + z_1 \\ &= m. \qquad (13) \end{aligned}$$

Therefore,

$$S_{z_1}(0) = m + R_{z_1}(0-m) = m + R_{z_1}(-m) = 2m = z_1.$$

Similarly, by linearity of R and (13),

$$S_{z_1}(z_1) = m + R_{z_1}(z_1 - m) = m + R_{z_1}(m) = m - R_{z_1}(-m) = 0.$$

5. We compute as follows:

$$\begin{aligned} \|S_{z_1}(y)\| &= \|S_{z_1}(y) - S_{z_1}(z_1)\| \qquad (14) \\ &= \|y - z_1\| \qquad (15) \end{aligned}$$

where we used part 4 of the lemma for (14), and part 3 of the lemma for (15). Similarly,

$$\begin{aligned} \|S_{z_1}(y) - z_1\| &= \|S_{z_1}(y) - S_{z_1}(0)\| \qquad (16) \\ &= \|y - 0\| \qquad (17) \\ &= \|y\| \end{aligned}$$

where we used part 4 of the lemma for (16), and part 3 of the lemma for (17). □

Lemma 4.8 (Constant identity II)**.** *Assume that Assumption **W.1** holds. Then,*

$$\sigma_\eta^{(k,p,2)} = \frac{1}{2}\,\sigma_\eta^{(k,p)}.$$

Proof. We define

$$\mathcal{Z} := \int_{(\mathbb{R}^d)^k} \widetilde{\eta}_p(z_1,\dots,z_k)\,\mathrm{d}z_1\dots\mathrm{d}z_k$$

and

$$\mathrm{d}\mathbb{Q}(z_1,\cdots,z_k) := \frac{1}{\mathcal{Z}}\widetilde{\eta}_p(z_1,\dots,z_k)\,\mathrm{d}z_1\dots\mathrm{d}z_k := f(z_1,\dots,z_k)\,\mathrm{d}z_1\dots\mathrm{d}z_k.$$

For $(Z_1,\dots,Z_k)\sim\mathbb{Q}$, we write $Z_1 = (X_1,\dots,X_d)^\top$ and $Y := (Z_2)_d$. Then,

$$\sigma_\eta^{(k,p)} = \mathcal{Z}\,\mathbb{E}[|X_d|^p] \qquad \text{and} \qquad \sigma_\eta^{(k,p,2)} = \mathcal{Z}\,\mathbb{E}[|X_d|^{p-2}X_dY]$$

where the expectation is taken with respect to $\mathbb{Q}$.

As shown in the proof of Lemma 4.5, since $\widetilde{\eta}_p$ depends only on norms and pairwise distances,

$$\widetilde{\eta}_p(Qz_1, \ldots, Qz_k) = \widetilde{\eta}_p(z_1, \ldots, z_k)$$

for all $Q \in O(d)$, where $O(d)$ denotes the group of orthogonal matrices in $\mathbb{R}^d$. We now fix $z_1 \neq 0$ and let

$$G_{z_1} := \{Q \in O(d) : Qz_1 = z_1\}$$

be the subgroup of orthogonal transformations fixing z_1. For $Q \in G_{z_1}$, we therefore have

$$\widetilde{\eta}_p(z_1, Qz_2, \ldots, Qz_k) = \widetilde{\eta}_p(z_1, z_2, \ldots, z_k). \tag{18}$$

The marginal density of Z_1 at z_1 is

$$f_{Z_1}(z_1) = \int_{(\mathbb{R}^d)^{k-1}} f(z_1, z_2, \ldots, z_k)\, \mathrm{d}z_2 \cdots \mathrm{d}z_k.$$

For $f_{Z_1}(z_1) > 0$, the conditional density of $(Z_2, \ldots, Z_k)$ given $Z_1 = z_1$ is

$$f_{Z_2,\ldots,Z_k|Z_1}(z_2, \ldots, z_k \mid z_1) = \frac{f(z_1, z_2, \ldots, z_k)}{f_{Z_1}(z_1)}.$$

For a fixed $Q \in G_{z_1}$, we consider the conditional density at the point $(Qz_2, \ldots, Qz_k)$:

$$\begin{aligned}
f_{Z_2,\ldots,Z_k|Z_1}(Qz_2, \ldots, Qz_k \mid z_1) &= \frac{f(z_1, Qz_2, \ldots, Qz_k)}{f_{Z_1}(z_1)} \\
&= \frac{\widetilde{\eta}_p(z_1, Qz_2, \ldots, Qz_k)}{\mathcal{Z} f_{Z_1}(z_1)} \\
&= f_{Z_2,\ldots,Z_k|Z_1}(z_2, \ldots, z_k \mid z_1) \qquad (19)
\end{aligned}$$

where we used (18) for (19). Thus, for any measurable set $A \subseteq (\mathbb{R}^d)^{k-1}$,

$$\begin{aligned}
\mathbb{P}\big((Z_2, \ldots, Z_k) \in A \mid Z_1 = z_1\big) &= \int_A f_{Z_2,\ldots,Z_k|Z_1}(z_2, \ldots, z_k \mid z_1)\, \mathrm{d}z_2 \cdots \mathrm{d}z_k \\
&= \int_A f_{Z_2,\ldots,Z_k|Z_1}(Qz_2, \ldots, Qz_k \mid z_1)\, \mathrm{d}z_2 \cdots \mathrm{d}z_k \\
&= \int_{QA} f_{Z_2,\ldots,Z_k|Z_1}(w_2, \ldots, w_k \mid z_1)\, \mathrm{d}w_2 \cdots \mathrm{d}w_k \qquad (20) \\
&= \mathbb{P}\big((Z_2, \ldots, Z_k) \in QA \mid Z_1 = z_1\big) \qquad (21)
\end{aligned}$$

where we used the change of variables $w_j = Qz_j$ for $2 \leq j \leq k$ and the fact that $|\det Q| = 1$ for (20). Equivalently, we obtain that

$$(Z_2, \ldots, Z_k) \mid (Z_1 = z_1) \overset{d}{=} (QZ_2, \ldots, QZ_k) \mid (Z_1 = z_1).$$

By picking $A = B \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d$ for $B \subseteq \mathbb{R}^d$ a measurable set, we obtain

$$\begin{aligned}
\mathbb{P}\big(Z_2 \in B \mid Z_1 = z_1\big) &= \mathbb{P}\big((Z_2, \ldots, Z_k) \in A \mid Z_1 = z_1\big) \\
&= \mathbb{P}\big((Z_2, \ldots, Z_k) \in QA \mid Z_1 = z_1\big) \qquad (22) \\
&= \mathbb{P}\big(Z_2 \in QB \mid Z_1 = z_1\big)
\end{aligned}$$

where we used (21) for (22), which implies that

$$Z_2 \mid (Z_1 = z_1) \overset{d}{=} QZ_2 \mid (Z_1 = z_1) \tag{23}$$

for all $Q \in G_{z_1}$. Taking expectations gives

$$\mathbb{E}[Z_2 \mid Z_1 = z_1] = Q\,\mathbb{E}[Z_2 \mid Z_1 = z_1]. \tag{24}$$

Specifically, this means that $\mathbb{E}[Z_2 \mid Z_1 = z_1]$ is a vector which is fixed by all $Q \in G_{z_1}$. We decompose $w := \mathbb{E}[Z_2 \mid Z_1 = z_1] = \left(w \cdot \frac{z_1}{\|z_1\|}\right)\frac{z_1}{\|z_1\|} + \left(w - \left(w \cdot \frac{z_1}{\|z_1\|}\right)\frac{z_1}{\|z_1\|}\right) := \left(w \cdot \frac{z_1}{\|z_1\|}\right)\frac{z_1}{\|z_1\|} + w_\perp$. Let $Q \in G_{z_1}$, then we compute as follows:

$$\begin{aligned} w &= Qw \qquad (25)\\ &= \left(w \cdot \frac{z_1}{\|z_1\|}\right)\frac{Qz_1}{\|z_1\|} + Qw_\perp \\ &= \left(w \cdot \frac{z_1}{\|z_1\|}\right)\frac{z_1}{\|z_1\|} + Qw_\perp \qquad (26)\\ &= w - w_\perp + Qw_\perp \end{aligned}$$

where we used (24) for (25) and the fact that $Qz_1 = z_1$ by assumption for (26). This implies that $w_\perp = Qw_\perp$ and, since the action of G_{z_1} on the orthogonal complement $z_1^\perp$ is the full orthogonal group $O(d-1)$ [60, Example 21.19], this implies that $w_\perp = 0$. We therefore conclude that

$$\mathbb{E}[Z_2 \mid Z_1 = z_1] = \alpha(z_1)\, z_1 \tag{27}$$

for some scalar $\alpha(z_1)$.

We recall that

$$\tilde{\eta}_{\mathrm{p}}(z_1, \ldots, z_k) = \left[\prod_{s=1}^{k} \eta(|z_s|)\right]\left[\prod_{j=2}^{k}\prod_{r=1}^{j-1} \eta(|z_j - z_r|)\right]$$

and let S_{z_1} be the map defined in Lemma 4.7. Now, we estimate as follows:

$$\begin{aligned} &\tilde{\eta}_{\mathrm{p}}(z_1, S_{z_1}(z_2), \ldots, S_{z_1}(z_k)) \\ &= \eta(|z_1|)\left[\prod_{s=2}^{k} \eta(|S_{z_1}(z_s)|)\right]\left[\prod_{j=2}^{k} \eta(|S_{z_1}(z_j) - z_1|)\right]\left[\prod_{j=2}^{k}\prod_{r=2}^{j-1} \eta(|S_{z_1}(z_j) - S_{z_1}(z_r)|)\right] \\ &= \eta(|z_1|)\left[\prod_{s=2}^{k} \eta(|S_{z_1}(z_s)|)\right]\left[\prod_{j=2}^{k} \eta(|S_{z_1}(z_j) - z_1|)\right]\left[\prod_{j=2}^{k}\prod_{r=2}^{j-1} \eta(|z_j - z_r|)\right] \qquad (28)\\ &= \eta(|z_1|)\left[\prod_{s=2}^{k} \eta(|z_s - z_1|)\right]\left[\prod_{j=2}^{k} \eta(|z_j|)\right]\left[\prod_{j=2}^{k}\prod_{r=2}^{j-1} \eta(|z_j - z_r|)\right] \qquad (29)\\ &= \tilde{\eta}_{\mathrm{p}}(z_1, z_2, \ldots, z_k) \qquad (30) \end{aligned}$$

where we used part 3 of Lemma 4.7 for (28) and part 4 of Lemma 4.7 for (29). This directly implies:

$$\begin{aligned} f_{Z_2,\ldots,Z_k|Z_1}(S_{z_1}(z_2), \ldots, S_{z_1}(z_k) \mid z_1) &= \frac{f(z_1, S_{z_1}(z_2), \ldots, S_{z_1}(z_k))}{f_{Z_1}(z_1)} \\ &= \frac{\tilde{\eta}_{\mathrm{p}}(z_1, S_{z_1}(z_2), \ldots, S_{z_1}(z_k))}{\mathcal{Z} f_{Z_1}(z_1)} \\ &= f_{Z_2,\ldots,Z_k|Z_1}(z_2, \ldots, z_k \mid z_1) \qquad (31) \end{aligned}$$

where we used (30) for (31). Therefore, for a measurable set $A \subseteq (\mathbb{R}^d)^{k-1}$,

$$\begin{aligned} \mathbb{P}((Z_2, \ldots, Z_k) \in A \,|\, Z_1 = z_1) &= \int_A f_{Z_2,\ldots,Z_k|Z_1}(z_2, \ldots, z_k \mid z_1)\,\mathrm{d}z_2 \cdots \mathrm{d}z_k \\ &= \int_A f_{Z_2,\ldots,Z_k|Z_1}(S_{z_1}(z_2), \ldots, S_{z_1}(z_k) \mid z_1)\,\mathrm{d}z_2 \cdots \mathrm{d}z_k \qquad (32) \end{aligned}$$

$$= \int_{S_{z_1}(A)} f_{Z_2,\dots,Z_k|Z_1}(w_2,\dots,w_k \mid z_1)\,\mathrm{d}w_2\cdots\mathrm{d}w_k \tag{33}$$

$$= \mathbb{P}((Z_2,\dots,Z_k) \in S_{z_1}(A) \mid Z_1 = z_1)$$

where we used (31) for (32), the change of variables $w_j = S_{z_1}(z_j)$ and the fact that $|\det DS_{z_1}| = 1$ (since S_{z_1} is an isometry by part 3 of Lemma 4.7) for (33). We conclude that

$$(Z_2,\dots,Z_k) \mid (Z_1 = z_1) \overset{d}{=} (S_{z_1}(Z_2),\dots,S_{z_1}(Z_k)) \mid (Z_1 = z_1),$$

and taking marginals, analogously to how we derived (23),

$$Z_2 \mid (Z_1 = z_1) \overset{d}{=} S_{z_1}(Z_2) \mid (Z_1 = z_1). \tag{34}$$

We now estimate as follows (and using the notation of Lemma 4.7):

$$\begin{aligned}
\alpha(z_1)z_1 &= \mathbb{E}[Z_2 \mid Z_1 = z_1] && (35)\\
&= \mathbb{E}[S_{z_1}(Z_2) \mid Z_1 = z_1] && (36)\\
&= \mathbb{E}[m + R_{z_1}(Z_2 - m) \mid Z_1 = z_1]\\
&= m + R_{z_1}(\mathbb{E}[Z_2 \mid Z_1 = z_1] - m) && (37)\\
&= m + R_{z_1}\left(\alpha(z_1)z_1 - \frac{z_1}{2}\right) && (38)\\
&= m + \left(\alpha(z_1) - \frac{1}{2}\right) R_{z_1}(z_1) && (39)\\
&= \frac{z_1}{2} + \left(\frac{1}{2} - \alpha(z_1)\right) z_1 && (40)\\
&= (1 - \alpha(z_1))z_1.
\end{aligned}$$

where we used (27) for (35), (34) for (36), part 1 of Lemma 4.7 for (37), (27) for (38), part 1 of Lemma 4.7 for (39) and the fact that $R_{z_1}(z_1) = -z_1$ for (40). Since $z_1 \neq 0$, we deduce that $\alpha(z_1) = \frac{1}{2}$ and

$$\mathbb{E}[Z_2 \mid Z_1 = z_1] = \frac{z_1}{2}$$

or equivalently

$$\mathbb{E}[(Z_2)_d \mid Z_1] = \mathbb{E}[Y \mid Z_1] = \frac{(Z_1)_d}{2} = \frac{X_d}{2}. \tag{41}$$

We conclude with the following computation:

$$\begin{aligned}
\mathbb{E}[|X_d|^{p-2}X_dY] &= \mathbb{E}[\mathbb{E}[|X_d|^{p-2}X_dY \mid Z_1]] && (42)\\
&= \mathbb{E}[|X_d|^{p-2}X_d\mathbb{E}[Y \mid Z_1]] && (43)\\
&= \frac{1}{2}\mathbb{E}[|X_d|^p]
\end{aligned}$$

where we use the tower property of conditional expectation for (42), the fact that X_d is measurable with respect to the σ-algebra induced by Z_1 and (41) for (43). From this, we directly obtain

$$\sigma_\eta^{(k,p,2)} = \mathcal{Z}\,\mathbb{E}[|X_d|^{p-2}X_dY] = \frac{1}{2}\,\mathcal{Z}\,\mathbb{E}[|X_d|^p] = \frac{1}{2}\,\sigma_\eta^{(k,p)}. \qquad \square$$

Corollary 4.9 (p-Laplacian)**.** *Assume that assumption **W.1** holds. Then,*

$$\begin{aligned}
\Delta_\infty^{(k,p)}(u)(x) = \Bigg(&\|\nabla u(x)\|_2^{p-2}\rho(x)^k\nabla\rho(x)\cdot\nabla u(x) \times \frac{2(\sigma_\eta^{(k,p)} + (k-1)\sigma_\eta^{(k,p,2)})}{(p-1)\sigma_\eta^{(k,p,1)}}\\
&+ \rho(x)^{k+1}\|\nabla u(x)\|_2^{p-2}\left[\Delta u(x) + \left(\frac{\sigma_\eta^{(k,p)}}{\sigma_\eta^{(k,p,1)}} - 1\right)\frac{\nabla u(x)^\top\nabla^2u(x)\nabla u(x)}{\|\nabla u(x)\|_2^2}\right]\Bigg)\frac{\sigma_\eta^{(k,p,1)}(p-1)}{2\rho(x)}
\end{aligned}$$

where Δ denotes the regular continuum Laplacian operator.

Proof. By Lemma 4.6, we have that

$$\frac{\sigma_\eta^{(k,p)}}{\sigma_\eta^{(k,p,1)}} - 1 = p - 2. \tag{44}$$

Similarly, we have

$$\begin{aligned}
\frac{2(\sigma_\eta^{(k,p)} + (k-1)\sigma_\eta^{(k,p,2)})}{(p-1)\sigma_\eta^{(k,p,1)}} &= \frac{2}{p-1}\left[\frac{\sigma_\eta^{(k,p)}}{\sigma_\eta^{(k,p,1)}} + (k-1)\frac{\sigma_\eta^{(k,p,2)}}{\sigma_\eta^{(k,p,1)}}\right] \\
&= \frac{2}{p-1}\left[(p-1) + (k-1)(p-1)\frac{\sigma_\eta^{(k,p,2)}}{\sigma_\eta^{(k,p)}}\right] \qquad (45) \\
&= \frac{2}{p-1}\left[(p-1) + \frac{(k-1)(p-1)}{2}\right] \qquad (46) \\
&= k+1 \qquad (47)
\end{aligned}$$

where we used Lemma 4.6 for (45) and Lemma 4.8 for (46). We then have:

$$\begin{aligned}
&\Bigg(\|\nabla u(x)\|_2^{p-2}\rho(x)^k \nabla\rho(x)\cdot\nabla u(x) \times \frac{2(\sigma_\eta^{(k,p)} + (k-1)\sigma_\eta^{(k,p,2)})}{(p-1)\sigma_\eta^{(k,p,1)}} \\
&+ \rho(x)^{k+1}\|\nabla u(x)\|_2^{p-2}\left[\Delta u(x) + \left(\frac{\sigma_\eta^{(k,p)}}{\sigma_\eta^{(k,p,1)}} - 1\right)\frac{\nabla u(x)^\top \nabla^2 u(x)\nabla u(x)}{\|\nabla u(x)\|_2^2}\right]\Bigg)\frac{\sigma_\eta^{(k,p,1)}(p-1)}{2\rho(x)} \\
&= \Bigg(\|\nabla u(x)\|_2^{p-2}\rho(x)^k\nabla\rho(x)\cdot\nabla u(x)(k+1) \\
&+ \rho(x)^{k+1}\|\nabla u(x)\|_2^{p-2}\left[\Delta u(x) + (p-2)\frac{\nabla u(x)^\top\nabla^2 u(x)\nabla u(x)}{\|\nabla u(x)\|_2^2}\right]\Bigg)\frac{\sigma_\eta^{(k,p,1)}(p-1)}{2\rho(x)} \qquad (48) \\
&= \frac{\sigma_\eta^{(k,p)}}{2\rho(x)}\mathrm{div}(\rho(x)^{k+1}\|\nabla u\|_2^{p-2}\nabla u(x)) \qquad (49)
\end{aligned}$$

where we used (44) and (47) for (48) and Lemma 4.6 for (49). □

4.2.3 Proof of Theorem 3.2

Proof of Theorem 3.2. In the proof $C > 0$ will denote a constant that can be arbitrarily large, independent of n and that may change from line to line. We will roughly follow the strategy in [15].

Let us start by assuming that $p \geq 3$. By Taylor's theorem, if $\psi(t) = |t|^{p-2}t$, then, we have that $\psi(t) = \psi(a) + \psi'(a)(t-a) + \mathcal{O}(C_b^{p-3}|t-a|^2)$ for $a, t \in [-C_b, C_b]$. For $u \in \mathrm{C}^3(\mathbb{R}^d)$, let $t = u(x+z) - u(x)$ and $a = \nabla u(x)\cdot z$. Then, using the previous identity, we obtain that

$$\begin{aligned}
&\psi(u(x+z) - u(x)) \\
&= |\nabla u(x)\cdot z|^{p-2}\nabla u(x)\cdot z + (p-1)|\nabla u(x)\cdot z|^{p-2}(u(x+z) - u(x) - \nabla u(x)\cdot z) \\
&+ \mathcal{O}(C_b^{p-3}|u(x+z) - u(x) - \nabla u(x)\cdot z|^2).
\end{aligned}$$

Noting that $u(x+z) - u(x) - \nabla u(x)\cdot z = z^\top\nabla^2 u(x) z/2 + \mathcal{O}(\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}|z|^3)$ and also $u(x+z) - u(x) - \nabla u(x)\cdot z = \mathcal{O}(\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}|z|^2)$, we continue the above computation:

$$\begin{aligned}
\psi(u(x+z) - u(x)) = |\nabla u(x)\cdot z|^{p-2}\nabla u(x)\cdot z &+ \frac{(p-1)}{2}|\nabla u(x)\cdot z|^{p-2}z^\top\nabla^2 u(x) z \\
&+ \mathcal{O}(\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}|z|^{p+1}) + \mathcal{O}(C_b^{p-3}\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^2|z|^4).
\end{aligned}$$

Finally, we note that $\max\{|u(x+z) - u(x)|, |\nabla u(x)\cdot z|\} \leq C_b$ means that we can pick $C_b = \|u\|_{\mathrm{C}^3(\mathbb{R}^d)}|z|$ which allows us to conclude that

$$\psi(u(x+z) - u(x)) = |\nabla u(x)\cdot z|^{p-2}\nabla u(x)\cdot z + \frac{(p-1)}{2}|\nabla u(x)\cdot z|^{p-2}z^\top\nabla^2 u(x) z$$

$$+\mathcal{O}(\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}|z|^{p+1}). \tag{50}$$

We first start by assuming that $x_{i_0} \in \Omega_n \cap \Omega'$ is fixed (and hence non-random) and let $1 \leq k \leq q$ be fixed. Let us estimate as follows:

$$\begin{aligned} n^k \varepsilon_n^{p+kd} \Delta_{n,\varepsilon_n}^{(k,p)}(u)(x_{i_0}) =& \sum_{i_1,\dots,i_k=1}^n \eta_{\mathrm{p}}(x_{i_0},\dots,x_{i_k}) |(x_{i_1}-x_{i_0})\cdot\nabla u(x_{i_0})|^{p-2}(x_{i_1}-x_{i_0})\cdot\nabla u(x_{i_0}) \\ &+ \frac{(p-1)}{2}\sum_{i_1,\dots,i_k=1}^n \eta_{\mathrm{p}}(x_{i_0},\dots,x_{i_k}) |(x_{i_1}-x_{i_0})\cdot\nabla u(x_{i_0})|^{p-2}(x_{i_1}-x_{i_0})^\top\nabla^2 u(x_{i_0})(x_{i_1}-x_{i_0}) \qquad (51) \\ &+ \sum_{i_1,\dots,i_k=1}^n \eta_{\mathrm{p}}(x_{i_0},\dots,x_{i_k}) \mathcal{O}\left(\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}|x_{i_1}-x_{i_0}|^{p+1}\right) \\ &=: T_1(x_1,\dots,x_n)+T_2(x_1,\dots,x_n)+T_3(x_1,\dots,x_n) \\ &= \mathbb{E}(T_1)+\mathbb{E}(T_2)+\mathbb{E}(T_3)+\sum_{i=1}^3 (T_i-\mathbb{E}(T_i)) \end{aligned}$$

where we used (50) with $x = x_{i_0}$ and $z = x_{i_1} - x_{i_0}$ for (51).

We now want to estimate $|T_i - \mathbb{E}(T_i)|$ for $1 \leq i \leq 3$ using Theorem 4.1. For the purpose of the next few equations, for a general function $f(x_{i_0},\dots,x_{i_k})$, we will write $f(x_{i_0},\dots,x_{i_k})|_{\{x_1,\dots,x_n\}}$ where the extra subscript $\{x_1,\dots,x_n\}$ indicates that $x_{i_\ell} \in \{x_1,\dots,x_n\}$ for $1 \leq \ell \leq k$. Let us start by considering

$$\begin{aligned} &|T_3(x_1,\dots,x_i,\dots,x_n) - T_3(x_1,\dots,\tilde{x}_i,\dots,x_n)| \\ &\leq \sum_{i_1,\dots,i_k=1}^n \Bigg| \left[\eta_{\mathrm{p}}(x_{i_0},\dots,x_{i_k})\mathcal{O}\left(\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}|x_{i_1}-x_{i_0}|^{p+1}\right)\right]|_{\{x_1,\dots,x_i,\dots,x_n\}} \\ &\quad - \left[\eta_{\mathrm{p}}(x_{i_0},\dots,x_{i_k})\mathcal{O}\left(\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}|x_{i_1}-x_{i_0}|^{p+1}\right)\right]|_{\{x_1,\dots,\tilde{x}_i,\dots,x_n\}}\Bigg|. \end{aligned}$$

We note that each term in the latter sum is different from 0 only if there exists $1 \leq \ell \leq k$ with $i_\ell = i$. By Lemma 4.2, there exists $S^{(n,k)}(i) \leq kn^{k-1}$ such cases and for each of those, the term in the sum can be bounded by $C\varepsilon^{p+1}\|\eta\|_{\mathrm{L}^\infty}^{t(k)}\|u\|_{\mathrm{C}^2(\mathbb{R}^d)}^{p-1}$ where $t(k)$ is the number of terms in the double product in η_{p}. By Assumptions **W.1** and **S.1**, this leads to:

$$|T_3(x_1,\dots,x_i,\dots,x_n) - T_3(x_1,\dots,\tilde{x}_i,\dots,x_n)| \leq Cn^{k-1}\varepsilon^{p+1}\|u\|_{\mathrm{C}^3}^{p-1}$$

Similarly,

$$|T_1(x_1,\dots,x_i,\dots,x_n) - T_1(x_1,\dots,\tilde{x}_i,\dots,x_n)| \leq Cn^{k-1}\|\eta\|_{\mathrm{L}^\infty}^{t(k)}\varepsilon^{p-1}\|u\|_{\mathrm{C}^1}^{p-1}$$

and

$$|T_2(x_1,\dots,x_i,\dots,x_n) - T_2(x_1,\dots,\tilde{x}_i,\dots,x_n)| \leq Cn^{k-1}\varepsilon^{p}\|\eta\|_{\mathrm{L}^\infty}^{t(k)}\|u\|_{\mathrm{C}^2}^{p-1}.$$

Using Theorem 4.1, we therefore obtain that $\mathbb{P}(|T_i - \mathbb{E}(T_i)| \geq t) \leq 2\exp\left(-\frac{t^2}{Cn^{2k-1}\varepsilon_n^{2p-2}\|u\|_{\mathrm{C}^3}}\right)$ and with $t = n^k\varepsilon_n^{p+kd}\delta\|u\|_{\mathrm{C}^3}^{p-1}$,

$$\mathbb{P}(|T_i - \mathbb{E}(T_i)| \geq n^k\varepsilon_n^{p+kd}\delta) \leq 2\exp\left(-Cn\varepsilon_n^{2(1+kd)}\delta^2\right)$$

for $1 \leq i \leq 3$.

We now estimate $\mathbb{E}(T_i)$ for $1 \leq i \leq 3$. In particular,

$$\frac{1}{n^k\varepsilon_n^{p+kd}}\mathbb{E}(T_3) = \frac{1}{n^k\varepsilon_n^{p+kd}}\mathcal{O}\left(\varepsilon_n^{p+1}\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\mathbb{E}\left(\sum_{i_1,\dots,i_k=1}^n \eta_{\mathrm{p}}(x_{i_0},\dots,x_{i_k})\right)\right) \tag{52}$$

$$= \mathcal{O}\left(\varepsilon_n \|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1} \frac{1}{\varepsilon_n^{kd}} \int_{\Omega^k} \eta_{\mathrm{p}}(x_{i_0}, x_1, \dots, x_k) \left[\prod_{\ell=1}^k \rho(x_\ell)\right] \mathrm{d}x_k \cdots \mathrm{d}x_1\right)$$

$$= \mathcal{O}\left(\varepsilon_n \|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1} \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(z_1, \dots, z_k) \, \mathrm{d}z_k \cdots \mathrm{d}z_1\right) \tag{53}$$

$$= \mathcal{O}\left(\varepsilon_n \|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right) \tag{54}$$

where we used Assumption **W.1** to deduce that $|x_{i_0} - x_{i_1}| = \mathcal{O}(\varepsilon_n)$ for (52), Assumption **M.2** and the change of variables $z_j = (x_j - x_{i_0})/\varepsilon_n$ for $1 \leq j \leq k$ for (53) as well as Assumption **W.1** for (54).

For T_1, for n large enough, we proceed as follows:

$$\frac{1}{n^k \varepsilon_n^{p+kd}} \mathbb{E}(T_1)$$
$$= \frac{1}{\varepsilon_n^{p+kd}} \int_{\Omega^k} \eta_{\mathrm{p}}(x_{i_0}, x_1, \dots, x_k)$$
$$\times |(x_1 - x_{i_0}) \cdot \nabla u(x_{i_0})|^{p-2} (x_1 - x_{i_0}) \cdot \nabla u(x_{i_0}) \left[\prod_{\ell=1}^k \rho(x_\ell)\right] \mathrm{d}x_k \cdots \mathrm{d}x_1$$
$$= \frac{1}{\varepsilon_n} \int_{\otimes_j(\{z_j \mid x_{i_0} + \varepsilon_n z_j \in \Omega\} \cap \mathrm{supp}(\eta))} \tilde{\eta}_{\mathrm{p}}(z_1, \dots, z_k) |z_1 \cdot \nabla u(x_{i_0})|^{p-2} z_1 \cdot \nabla u(x_{i_0})$$
$$\times \left[\prod_{\ell=1}^k \rho(x_{i_0} + \varepsilon_n z_\ell)\right] \mathrm{d}z_k \cdots \mathrm{d}z_1 \tag{55}$$
$$= \frac{1}{\varepsilon_n} \int_{\mathrm{supp}(\eta)^k} \tilde{\eta}_{\mathrm{p}}(z_1, \dots, z_k) |z_1 \cdot \nabla u(x_{i_0})|^{p-2} z_1 \cdot \nabla u(x_{i_0}) \rho(x_{i_0})^k \, \mathrm{d}z_k \cdots \mathrm{d}z_1$$
$$+ \int_{\mathrm{supp}(\eta)^k} \tilde{\eta}_{\mathrm{p}}(z_1, \dots, z_k) |z_1 \cdot \nabla u(x_{i_0})|^{p-2} z_1 \cdot \nabla u(x_{i_0}) \rho(x_{i_0})^{k-1} \nabla \rho(x_{i_0})(z_1 + \dots + z_k) \, \mathrm{d}z_k \cdots \mathrm{d}z_1 \tag{56}$$
$$+ \mathcal{O}\left(\varepsilon_n \int_{\mathrm{supp}(\eta)^k} \tilde{\eta}_{\mathrm{p}}(z_1, \dots, z_k) |z_1 \cdot \nabla u(x_{i_0})|^{p-2} z_1 \cdot \nabla u(x_{i_0}) \, \mathrm{d}z_k \cdots \mathrm{d}z_1\right)$$
$$= \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(z_1, \dots, z_k) |z_1 \cdot \nabla u(x_{i_0})|^{p-2} z_1 \cdot \nabla u(x_{i_0}) \rho(x_{i_0})^{k-1} \nabla \rho(x_{i_0})(z_1 + \dots + z_k) \, \mathrm{d}z_k \cdots \mathrm{d}z_1$$
$$+ \mathcal{O}\left(\varepsilon_n \|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right) \tag{57}$$
$$= \rho(x_{i_0})^{k-1} \sum_{i=1}^d \frac{\partial \rho}{\partial x_i}(x_{i_0}) \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(z_1, \dots, z_k) \psi\left(z_1 \cdot \nabla u(x_{i_0})\right) (z_1 + \dots + z_k)_i \, \mathrm{d}z_k \cdots \mathrm{d}z_1$$
$$+ \mathcal{O}\left(\varepsilon_n \|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right) \tag{58}$$

where we used the change of variables $z_j = (x_j - x_{i_0})/\varepsilon_n$ for $1 \leq j \leq k$ for (55), Lemmas 4.3 and 4.4 for (56), Assumption **W.1** as well as the fact that $f(z_1, \dots, z_k) := \tilde{\eta}_{\mathrm{p}}(z_1, \dots, z_k) |z_1 \cdot \nabla u(x_{i_0})|^{p-2} z_1 \cdot \nabla u(x_{i_0}) \rho(x_{i_0})^k$ is odd symmetric for (57), and recalling $\psi(t) = |t|^{p-2} t$. Let O be the orthogonal matrix so that $O e_d = \nabla u(x_{i_0}) / \|\nabla u(x_{i_0})\|_2$ where $e_d = (0, \dots, 0, 1) \in \mathbb{R}^d$. By the change of variables $\tilde{z}_j = O^\top z_j$ for $1 \leq j \leq k$ and noting that $z_1 \cdot \nabla u(x_{i_0}) = \tilde{z}_1 \cdot O^\top \nabla u(x_{i_0}) = (\tilde{z}_1)_d \|\nabla u(x_{i_0})\|_2$, we can continue our computation from (58):

$$\frac{1}{n^k \varepsilon_n^{p+kd}} \mathbb{E}(T_1)$$
$$= \|\nabla u(x_{i_0})\|_2^{p-1} \rho(x_{i_0})^{k-1} \sum_{i=1}^d \frac{\partial \rho}{\partial x_i}(x_{i_0}) \left[\int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1, \dots, \tilde{z}_k) \psi\left((\tilde{z}_1)_d\right)\right.$$
$$\left. \times \sum_{j=1}^d (O)_{ij} (\tilde{z}_1 + \dots + \tilde{z}_k)_j \, \mathrm{d}\tilde{z}_k \cdots \mathrm{d}\tilde{z}_1\right] + \mathcal{O}\left(\varepsilon_n \|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right)$$

$$= \|\nabla u(x_{i_0})\|_2^{p-1}\rho(x_{i_0})^{k-1}\sum_{i,j=1}^d \frac{\partial \rho}{\partial x_i}(x_{i_0})(O)_{ij}\sum_{r=1}^k \int_{(\mathbb{R}^d)^k}\tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)\psi\left((\tilde{z}_1)_d\right)(\tilde{z}_r)_j \,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1$$
$$+\,\mathcal{O}\left(\varepsilon_n\|u\|^{p-1}_{\mathrm{C}^3(\mathbb{R}^d)}\right).$$

For $j \neq d$, we note that

$$T_4 := \int_{(\mathbb{R}^d)^k}\tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)\psi\left((\tilde{z}_1)_d\right)(\tilde{z}_r)_j \,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1$$
$$= \int_{(\mathbb{R})^{k(d-1)}}(\tilde{z}_r)_j\left[\int_{\mathbb{R}^k}\tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)\psi\left((\tilde{z}_1)_d\right)\mathrm{d}(\tilde{z}_k)_d\cdots\mathrm{d}(\tilde{z}_1)_d\right]\mathrm{d}(\tilde{z}_k)_{1:d-1}\cdots\mathrm{d}(\tilde{z}_1)_{1:d-1}$$

(denoting by $(a)_{j:k}$ the elements $a_j, a_{j+1}\dots, a_{k-1}, a_k$) and the function $f:\mathbb{R}^k \mapsto \mathbb{R}$ defined as

$$f(y_1,\dots,y_k) = \tilde{\eta}_{\mathrm{p}}((\tilde{z}_1)_{1:d-1}, y_1, (\tilde{z}_2)_{1:d-1}, y_2,\dots,(\tilde{z}_k)_{1:d-1}, y_k)\psi\left(y_1\right)$$

is odd symmetric for any fixed $(\tilde{z}_1)_{1:d-1},\dots,(\tilde{z}_k)_{1:d-1}$ and therefore $T_4 = 0$ and

$$\frac{1}{n^k\varepsilon_n^{p+kd}}\mathbb{E}(T_1)$$
$$= \|\nabla u(x_{i_0})\|_2^{p-1}\rho(x_{i_0})^{k-1}\sum_{i=1}^d \frac{\partial \rho}{\partial x_i}(x_{i_0})(O)_{id}\sum_{r=1}^k \int_{(\mathbb{R}^d)^k}\tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)\psi((\tilde{z}_1)_d)(\tilde{z}_r)_d \,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1$$
$$+\,\mathcal{O}\left(\varepsilon_n\|u\|^{p-1}_{\mathrm{C}^3(\mathbb{R}^d)}\right)$$
$$= \|\nabla u(x_{i_0})\|_2^{p-1}\rho(x_{i_0})^{k-1}\sum_{i=1}^d \frac{\partial \rho}{\partial x_i}(x_{i_0})(O)_{id}(\sigma_\eta^{(k,p)} + (k-1)\sigma_\eta^{(k,p,2)}) + \mathcal{O}\left(\varepsilon_n\|u\|^{p-1}_{\mathrm{C}^3(\mathbb{R}^d)}\right).$$

By recalling that $(O)_{id} = (\nabla u(x_{i_0}))_i/\|\nabla u(x_{i_0})\|_2$, we can conclude:

$$\frac{1}{n^k\varepsilon_n^{p+kd}}\mathbb{E}(T_1)$$
$$= \|\nabla u(x_{i_0})\|_2^{p-2}\rho(x_{i_0})^{k-1}\nabla\rho(x_{i_0})\cdot\nabla u(x_{i_0})(\sigma_\eta^{(k,p)} + (k-1)\sigma_\eta^{(k,p,2)}) + \mathcal{O}\left(\varepsilon_n\|u\|^{p-1}_{\mathrm{C}^3(\mathbb{R}^d)}\right). \tag{59}$$

Let us now tackle T_2. We can estimate as follows, for n large enough:

$$\frac{1}{n^k\varepsilon_n^{p+kd}}\mathbb{E}(T_2)$$
$$= \frac{(p-1)}{2}\int_{(\mathbb{R}^d)^k}\tilde{\eta}_{\mathrm{p}}(z_1,\dots,z_k)|z_1\cdot\nabla u(x_{i_0})|^{p-2}z_1^\top\nabla^2 u(x_{i_0})z_1\left[\prod_{\ell=1}^k\rho(x_{i_0}+\varepsilon_n z_\ell)\right]\mathrm{d}z_k\cdots\mathrm{d}z_1 \tag{60}$$
$$= \frac{(p-1)}{2}\rho(x_{i_0})^k\int_{(\mathbb{R}^d)^k}\tilde{\eta}_{\mathrm{p}}(z_1,\dots,z_k)|z_1\cdot\nabla u(x_{i_0})|^{p-2}z_1^\top\nabla^2 u(x_{i_0})z_1\,\mathrm{d}z_k\cdots\mathrm{d}z_1$$
$$+\,\mathcal{O}\left(\varepsilon_n\|u\|^{p-1}_{\mathrm{C}^3(\mathbb{R}^d)}\right) \tag{61}$$
$$= \frac{(p-1)}{2}\rho(x_{i_0})^k\|\nabla u(x_{i_0})\|_2^{p-2}\int_{(\mathbb{R}^d)^k}\tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)|(\tilde{z}_1)_d|^{p-2}(O\tilde{z}_1)^\top\nabla^2 u(x_{i_0})(O\tilde{z}_1)\,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1$$
$$+\,\mathcal{O}\left(\varepsilon_n\|u\|^{p-1}_{\mathrm{C}^3(\mathbb{R}^d)}\right) \tag{62}$$
$$= \frac{(p-1)}{2}\rho(x_{i_0})^k\|\nabla u(x_{i_0})\|_2^{p-2}\int_{(\mathbb{R}^d)^k}\tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)|(\tilde{z}_1)_d|^{p-2}$$
$$\times\sum_{i,j=1}^d(\nabla^2 u(x_{i_0}))_{ij}(O\tilde{z}_1)_i(O\tilde{z}_1)_j\,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1 + \mathcal{O}\left(\varepsilon_n\|u\|^{p-1}_{\mathrm{C}^3(\mathbb{R}^d)}\right)$$

$$
\begin{aligned}
&= \frac{(p-1)}{2}\rho(x_{i_0})^k \|\nabla u(x_{i_0})\|_2^{p-2} \sum_{i,j=1}^{d} (\nabla^2 u(x_{i_0}))_{ij} \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)|(\tilde{z}_1)_d|^{p-2} \\
&\times \sum_{r,\ell=1}^{d} (O)_{ir}(O)_{j\ell}(\tilde{z}_1)_\ell(\tilde{z}_1)_r \,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1 + \mathcal{O}\left(\varepsilon_n\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right) \\
&= \frac{(p-1)}{2}\rho(x_{i_0})^k \|\nabla u(x_{i_0})\|_2^{p-2} \sum_{i,j=1}^{d} (\nabla^2 u(x_{i_0}))_{ij} \sum_{r,\ell=1}^{d} (O)_{ir}(O)_{j\ell} \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)|(\tilde{z}_1)_d|^{p-2} \\
&\times (\tilde{z}_1)_\ell(\tilde{z}_1)_r \,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1 + \mathcal{O}\left(\varepsilon_n\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right)
\end{aligned}
$$

where we used the change of variables $z_j = (x_j - x_{i_0})/\varepsilon_n$ for $1 \le j \le k$ and Lemma 4.4 for (60), Lemma 4.3 and Assumption **W.1** for (61) and the change of variables $\tilde{z}_j = O^\top z_j$ for $1 \le j \le k$ for (62). Similarly to the above, for $\ell \neq r \neq d$,

$$
\begin{aligned}
T_5 &:= \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)|(\tilde{z}_1)_d|^{p-2}(\tilde{z}_1)_\ell(\tilde{z}_1)_r \,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1 \\
&= \int_{(\mathbb{R})^{k(d-1)}} |(\tilde{z}_1)_d|^{p-2}(\tilde{z}_1)_\ell \left[\int_{\mathbb{R}^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)(\tilde{z}_1)_r \,\mathrm{d}(\tilde{z}_k)_r\cdots\mathrm{d}(\tilde{z}_1)_r\right] \mathrm{d}(\tilde{z}_k)_{-r}\cdots\mathrm{d}(\tilde{z}_1)_{-r}
\end{aligned}
$$

(denoting by $(a)_{-r}$ the vector $(a_1,\dots,a_{r-1},a_{r+1},\dots,a_d)$) and the function $f:\mathbb{R}^k \mapsto \mathbb{R}$ defined as

$$
f(y_1,\dots,y_k) = \tilde{\eta}_{\mathrm{p}}((\tilde{z}_1)_{1:r-1}, y_1, (\tilde{z}_1)_{r+1:d}, (\tilde{z}_2)_{1:r-1}, y_2, (\tilde{z}_2)_{r+1:d}, \dots, (\tilde{z}_k)_{r+1:d})y_1
$$

is odd symmetric for any fixed $(\tilde{z}_1)_{-r},\dots,(\tilde{z}_k)_{-r}$, so $T_5 = 0$ in this case. By symmetry the case $r \neq \ell \neq d$ follows. We therefore have:

$$
\begin{aligned}
&\frac{1}{n^k\varepsilon_n^{p+kd}}\mathbb{E}(T_2) \\
&= \frac{(p-1)}{2}\rho(x_{i_0})^k \|\nabla u(x_{i_0})\|_2^{p-2} \sum_{i,j=1}^{d} (\nabla^2 u(x_{i_0}))_{ij} \sum_{r=1}^{d} (O)_{ir}(O)_{jr} \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)|(\tilde{z}_1)_d|^{p-2} \\
&\times (\tilde{z}_1)_r^2 \,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1 + \mathcal{O}\left(\varepsilon_n\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right) \\
&= \frac{(p-1)}{2}\rho(x_{i_0})^k \|\nabla u(x_{i_0})\|_2^{p-2} \sum_{r=1}^{d} (O^\top \nabla^2 u(x_{i_0}) O)_{rr} \int_{(\mathbb{R}^d)^k} \tilde{\eta}_{\mathrm{p}}(\tilde{z}_1,\dots,\tilde{z}_k)|(\tilde{z}_1)_d|^{p-2} \\
&\times (\tilde{z}_1)_r^2 \,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1 + \mathcal{O}\left(\varepsilon_n\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right) \\
&= \frac{(p-1)}{2}\rho(x_{i_0})^k \|\nabla u(x_{i_0})\|_2^{p-2} \\
&\times \left[\mathrm{Tr}(\nabla^2 u(x_{i_0}))\sigma_\eta^{(k,p,1)} + (\sigma_\eta^{(k,p)} - \sigma_\eta^{(k,p,1)})(O^\top \nabla^2 u(x_{i_0}) O)_{dd}\right] \\
&+ \mathcal{O}\left(\varepsilon_n\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right) \\
&= \frac{(p-1)}{2}\rho(x_{i_0})^k \|\nabla u(x_{i_0})\|_2^{p-2} \Big[\mathrm{Tr}(\nabla^2 u(x_{i_0}))\sigma_\eta^{(k,p,1)} \\
&+ (\sigma_\eta^{(k,p)} - \sigma_\eta^{(k,p,1)}) \frac{1}{\|\nabla u(x_{i_0})\|_2^2} \nabla u(x_{i_0})^\top \nabla^2 u(x_{i_0}) \nabla u(x_{i_0})\Big] + \mathcal{O}\left(\varepsilon_n\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right) \qquad (63)
\end{aligned}
$$

where we used the fact that

$$
(O^\top \nabla^2 u(x_{i_0}) O)_{dd} = \sum_{i,j=1}^{n} (\nabla^2 u(x_{i_0}))_{ij}(O)_{id}(O)_{jd} = \sum_{i,j=1}^{n} (\nabla^2 u(x_{i_0}))_{ij} \frac{(\nabla u(x_{i_0}))_i}{\|\nabla u(x_{i_0})\|_2} \frac{(\nabla u(x_{i_0}))_j}{\|\nabla u(x_{i_0})\|_2}
$$

for (63).

Combining (59), (63) and (54), we obtain that, with probability $1-6\exp\left(-Cn\varepsilon_n^{2(1+kd)}\delta^2\right)$, since $\varepsilon_n \le \delta$:

$$\begin{aligned}\Delta_{n,\varepsilon_n}^{(k,p)}(u)(x_{i_0}) &= \|\nabla u(x_{i_0})\|_2^{p-2}\rho(x_{i_0})^{k-1}\nabla\rho(x_{i_0})\cdot\nabla u(x_{i_0})(\sigma_\eta^{(k,p)}+(k-1)\sigma_\eta^{(k,p,2)})\\ &+\frac{(p-1)}{2}\rho(x_{i_0})^k\|\nabla u(x_{i_0})\|_2^{p-2}\Big[\mathrm{Tr}(\nabla^2 u(x_{i_0}))\sigma_\eta^{(k,p,1)}\\ &+(\sigma_\eta^{(k,p)}-\sigma_\eta^{(k,p,1)})\frac{1}{\|\nabla u(x_{i_0})\|_2^2}\nabla u(x_{i_0})^\top\nabla^2 u(x_{i_0})\nabla u(x_{i_0})\Big]\\ &+\mathcal{O}\left(\delta\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right).\end{aligned}$$

As in [15], by taking a union bound on all $x_{i_0}\in\Omega_n\cap\Omega'$ and using Corollary 4.9, we obtain that

$$\left|\Delta_{n,\varepsilon_n}^{(k,p)}(u)(x_{i_0})-\rho(x_{i_0})\Delta_\infty^{(k,p)}(u)(x_{i_0})\right|\le\mathcal{O}\left(\delta\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}^{p-1}\right)$$

with probability $1-Cn\exp\left(-Cn\varepsilon_n^{2(1+kd)}\delta^2\right)$. To conclude the proof, we sum over $1\le k\le q$.

When $p=2$, we have $\psi(t)=t$ and directly obtain the estimate (50):

$$u(x+z)-u(x)=\nabla u(x)\cdot z+\frac{1}{2}z^\top\nabla^2 u(x)z+\mathcal{O}(\|u\|_{\mathrm{C}^3(\mathbb{R}^d)}|z|^3).$$

For the remainder of the proof, we proceed exactly as above with p replaced by 2. □

4.3 Γ-convergence

By re-adapting the results in [39,84], we are able to show the following Γ-convergence results. In particular, we note that we perform the following decomposition of our problem: we first show Γ-convergence of a nonlocal version of our continuum energies to their local counterparts; next, we establish Γ-convergence of the discrete energies to the nonlocal continuum energies.

We will use the following inequality often in our computations. For $a,b\in\mathbb{R}$, $\delta>0$ and $p>1$, there exists a constant C_δ such that

$$||c|^p-|a|^p|\le C_\delta|c-a|^p+\delta|a|^p. \tag{64}$$

We also note that $C_\delta\to\infty$ as $\delta\to 0$.

4.3.1 Γ-convergence of the nonlocal energies

For $v:\Omega\mapsto\mathbb{R}$ and $\varepsilon>0$ we define the nonlocal energies

$$\mathcal{E}_{\varepsilon,\mathrm{NL}}^{(k,p)}(v,\eta)=\frac{1}{\varepsilon^{p+kd}}\int_{\Omega^{k+1}}\left[\prod_{j=1}^{k}\prod_{r=0}^{j-1}\eta\left(\frac{|x_j-x_r|}{\varepsilon}\right)\right]|v(x_1)-v(x_0)|^p\prod_{\ell=0}^{k}\rho(x_\ell)\,\mathrm{d}x_k\cdots\mathrm{d}x_0$$

which are useful intermediary quantities when going from the discrete setting to the continuum one. In this Section, by re-adapting the results in [39], our aim is to prove the Γ-convergence of our nonlocal energies to the local ones in the continuum. We start with a few technical lemmas used in the subsequent results.

Lemma 4.10 (Integral identity)**.** *Assume that* ***S.1*** *and* ***W.1*** *hold. For* $k\ge 1$*, we have*

$$\frac{1}{\varepsilon_n^{p+dk}}\int_{\Omega^{k+1}}\left[\prod_{j=1}^{k}\prod_{r=0}^{j-1}\eta\left(\frac{|x_j-x_r|}{\varepsilon_n}\right)\right]|x_1-x_0|^{2p}\,\mathrm{d}x_k\cdots\mathrm{d}x_0=O(\varepsilon_n^p).$$

Proof. In the proof $C>0$ will denote a constant that can be arbitrarily large, is independent of n and that may change from line to line.

By using the change of variables $z_j = (x_j - x_0)/\varepsilon_n$ for $1 \le j \le k$, we obtain that $(x_j - x_r)/\varepsilon_n = z_j - z_r$ for $1 \le r < j \le k$. By the latter,

$$
\begin{aligned}
T_1 &:= \frac{1}{\varepsilon_n^{p+dk}} \int_{\Omega^{k+1}} \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\varepsilon_n}\right)\right] |x_1 - x_0|^{2p} \,\mathrm{d}x_k \cdots \mathrm{d}x_0 \\
&= \frac{\varepsilon_n^{2p}}{\varepsilon_n^{p}} \int_{\Omega} \int_{\{z_j \mid x_0 + \varepsilon_n z_j \in \Omega\}} |z_1|^{2p} \left[\prod_{s=1}^{k} \eta(|z_s|)\right] \left[\prod_{j=1}^{k}\prod_{r=1}^{j-1} \eta\left(|z_j - z_r|\right)\right] \,\mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0 \\
&\le C\varepsilon_n^{p} \int_{(\mathbb{R}^d)^k} |z_1|^{2p} \left[\prod_{s=1}^{k} \eta(|z_s|)\right] \left[\prod_{j=1}^{k}\prod_{r=1}^{j-1} \eta\left(|z_j - z_r|\right)\right] \,\mathrm{d}z_k \cdots \mathrm{d}z_1 \\
&= O(\varepsilon_n^p)
\end{aligned}
$$

where the last equality follows from Assumption **W.1**. □

Lemma 4.11 (Product identities)**.** *Let $\rho : \mathbb{R}^d \mapsto \mathbb{R}$ be a Lipschitz function that is bounded above. For $x_0, z_1, \cdots, z_k \in \mathbb{R}^d$ and $k \ge 1$, we have the following identities:*

$$
\left|\prod_{r=1}^{k} \rho\left(z_r + x_0\right) - \rho(x_0)^k\right| \le C(\rho) \sum_{r=1}^{k} |z_r| \tag{65}
$$

and

$$
\left|\prod_{r=0}^{k} \rho\left(x_r + z\right) - \prod_{r=0}^{k} \rho(x_r)\right| \le C(\rho)\,|z| \tag{66}
$$

for constants $C(\rho)$ only depending on ρ.

Proof. We only show how to derive (65) as the proof of (66) is similar.

We proceed by induction. For $k = 1$, $|\rho(x_0 + z_1) - \rho(x_0)| \le \mathrm{Lip}(\rho)|z_1|$. Now assume that (65) holds for $k - 1$. We compute as follows:

$$
\begin{aligned}
\left|\prod_{r=1}^{k} \rho\left(z_r + x_0\right) - \rho(x_0)^k\right| &\le \left|\prod_{r=1}^{k} \rho\left(z_r + x_0\right) - \rho(x_0)\prod_{r=1}^{k-1} \rho\left(z_r + x_0\right)\right| + \left|\rho(x_0)\prod_{r=1}^{k-1} \rho\left(z_r + x_0\right) - \rho(x_0)^k\right| \\
&= \left|\prod_{r=1}^{k-1} \rho\left(z_r + x_0\right)\right| \left|\rho\left(z_k + x_0\right) - \rho(x_0)\right| \\
&+ |\rho(x_0)| \left|\prod_{r=1}^{k-1} \rho\left(z_r + x_0\right) - \rho(x_0)^{k-1}\right| \\
&\le \|\rho\|_{\mathrm{L}^\infty}^{k-1} \mathrm{Lip}(\rho)\,|z_k| + \|\rho\|_{\mathrm{L}^\infty} C(\rho) \sum_{r=1}^{k-1} |z_r|.
\end{aligned}
$$
□

Lemma 4.12 (Pointwise convergence of nonlocal energies)**.** *Assume that **S.1**, **M.1**, **M.2** and **W.1** hold. Let $\{v_{\varepsilon_n}\}$ be a sequence of functions in $\mathrm{C}^2(\mathbb{R}^d)$ such that*

$$
\sup_{n\in\mathbb{N}} \{\|\nabla v_{\varepsilon_n}\|_{\mathrm{L}^\infty(\mathbb{R}^d)} + \|\nabla^2 v_{\varepsilon_n}\|_{\mathrm{L}^\infty(\mathbb{R}^d)}\} < \infty. \tag{67}
$$

Suppose that ρ is a positive Lipschitz function and that $\nabla v_{\varepsilon_n} \to \nabla v^*$ in $\mathrm{L}^p(\Omega)$ for some $v^* \in \mathrm{C}^2(\mathbb{R}^d)$. Then,

$$
\lim_{n\to\infty} \mathcal{E}_{\varepsilon_n,\mathrm{NL}}^{(k,p)}(v_{\varepsilon_n}, \eta) = \mathcal{E}_\infty^{(k,p)}(v^*). \tag{68}
$$

Proof. In the proof $C>0$ will denote a constant that can be arbitrarily large, is independent of n and that may change from line to line.

For a function $v \in \mathrm{C}^2(\mathbb{R}^d)$ and $x_0, x_1 \in \Omega$, we have:

$$v(x_1) - v(x_0) = \nabla v(x_0)\cdot(x_1-x_0) + (x_1-x_0)^T \nabla^2 v(c)(x_1-x_0).$$

for some constant c depending on x_0 and x_1. Now, define

$$H_{\varepsilon_n}(v) = \frac{1}{\varepsilon_n^{p+dk}} \int_{\Omega^{k+1}} \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|x_j-x_r|}{\varepsilon_n}\right)\right] |\nabla v(x_0)\cdot(x_1-x_0)|^p \left[\prod_{\ell=0}^{k}\rho(x_\ell)\right] \mathrm{d}x_k\cdots\mathrm{d}x_0.$$

We note that by Assumption **W.1**, we have $H_{\varepsilon_n}(v) \le C\|\nabla v\|_{\mathrm{L}^\infty}$. Then, we estimate as follows for $\delta>0$:

$$\begin{aligned}
T_1 &:= |\mathcal{E}^{(k,p)}_{\varepsilon_n,\mathrm{NL}}(v_{\varepsilon_n},\eta) - H_{\varepsilon_n}(v_{\varepsilon_n})| \\
&\le \frac{CC_\delta}{\varepsilon_n^{p+dk}} \int_{\Omega^{k+1}} \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|x_j-x_r|}{\varepsilon_n}\right)\right] |v_{\varepsilon_n}(x_1) - v_{\varepsilon_n}(x_0) - \nabla v_{\varepsilon_n}(x_0)\cdot(x_1-x_0)|^p \,\mathrm{d}x_k\cdots\mathrm{d}x_0 \\
&\quad + \delta H_{\varepsilon_n}(v_{\varepsilon_n}) \qquad (69)\\
&\le \frac{C_\delta\|\nabla^2 v_{\varepsilon_n}\|_{\mathrm{L}^\infty(\mathbb{R}^d)}}{\varepsilon_n^{p+dk}} \int_{\Omega^{k+1}} \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|x_j-x_r|}{\varepsilon_n}\right)\right] |x_1-x_0|^{2p}\,\mathrm{d}x_k\cdots\mathrm{d}x_0 + \delta H_{\varepsilon_n}(v_{\varepsilon_n}) \\
&= C_\delta O(\varepsilon_n^p) + \delta H_{\varepsilon_n}(v_{\varepsilon_n}) \qquad (70)
\end{aligned}$$

where we used (64) and Assumption **M.2** for (69) as well as Lemma 4.10 for (70).

Next, we define

$$\begin{aligned}
\tilde{H}_{\varepsilon_n}(v) = \frac{1}{\varepsilon_n^{p+dk}} \int_\Omega \int_{\substack{\{z_j \,|\, x_0+z_j\in\Omega\} \\ \text{for all } 1\le j\le k}} \left[\prod_{s=1}^{k}\eta\left(\frac{|z_s|}{\varepsilon_n}\right)\right] \left[\prod_{j=1}^{k}\prod_{r=1}^{j-1} \eta\left(\frac{|z_j-z_r|}{\varepsilon_n}\right)\right] |\nabla v(x_0)\cdot z_1|^p \\
\times \rho(x_0)^{k+1}\,\mathrm{d}z_k\cdots\mathrm{d}z_1\mathrm{d}x_0.
\end{aligned}$$

We note that with the change of variables $z_j = x_j - x_0$ for $1\le j\le k$, we have

$$\begin{aligned}
H_{\varepsilon_n}(v) = \frac{1}{\varepsilon_n^{p+dk}} \int_\Omega \int_{\substack{\{z_j \,|\, x_0+z_j\in\Omega\} \\ \text{for all } 1\le j\le k}} \left[\prod_{s=1}^{k}\eta\left(\frac{|z_s|}{\varepsilon_n}\right)\right] \left[\prod_{j=1}^{k}\prod_{r=1}^{j-1} \eta\left(\frac{|z_j-z_r|}{\varepsilon_n}\right)\right] |\nabla v(x_0)\cdot z_1|^p \\
\times \rho(x_0)\prod_{t=1}^{k}\rho\left(z_t+x_0\right)\,\mathrm{d}z_k\cdots\mathrm{d}z_1\mathrm{d}x_0.
\end{aligned}$$

This leads us to

$$\begin{aligned}
&|H_{\varepsilon_n}(v_{\varepsilon_n}) - \tilde{H}_{\varepsilon_n}(v_{\varepsilon_n})| \le \frac{C\|\nabla v_{\varepsilon_n}\|_{\mathrm{L}^\infty}\|\rho\|_{\mathrm{L}^\infty}}{\varepsilon_n^{p+dk}} \int_\Omega \int_{\substack{\{z_j \,|\, x_0+z_j\in\Omega\} \\ \text{for all } 1\le j\le k}} |z_1|^p \\
&\times \left[\prod_{s=1}^{k}\eta\left(\frac{|z_s|}{\varepsilon_n}\right)\right] \left[\prod_{j=1}^{k}\prod_{r=1}^{j-1} \eta\left(\frac{|z_j-z_r|}{\varepsilon_n}\right)\right] \cdot \left|\prod_{t=1}^{k}\rho\left(z_t+x_0\right) - \rho(x_0)^k\right| \mathrm{d}z_k\cdots\mathrm{d}z_1\mathrm{d}x_0 \\
&\le \frac{C}{\varepsilon_n^{p+dk}} \int_\Omega \int_{\substack{\{z_j \,|\, x_0+z_j\in\Omega\} \\ \text{for all } 1\le j\le k}} |z_1|^p \cdot \left[\prod_{s=1}^{k}\eta\left(\frac{|z_s|}{\varepsilon_n}\right)\right] \left[\prod_{j=1}^{k}\prod_{r=1}^{j-1} \eta\left(\frac{|z_j-z_r|}{\varepsilon_n}\right)\right] \cdot \sum_{r=1}^{k}|z_r|\,\mathrm{d}z_k\cdots\mathrm{d}z_1\mathrm{d}x_0 \qquad (71)\\
&\le C\varepsilon_n \int_\Omega \int_{\substack{\{\tilde{z}_j \,|\, \varepsilon_n|\tilde{z}_j|\le \mathrm{diam}(\Omega)\} \\ \text{for all } 1\le j\le k}} |\tilde{z}_1|^p \cdot \left[\prod_{s=1}^{k}\eta\left(|\tilde{z}_s|\right)\right] \left[\prod_{j=1}^{k}\prod_{r=1}^{j-1} \eta\left(|\tilde{z}_j-\tilde{z}_r|\right)\right] \cdot \sum_{r=1}^{k}|\tilde{z}_r|\,\mathrm{d}\tilde{z}_k\cdots\mathrm{d}\tilde{z}_1\mathrm{d}x_0 \qquad (72)\\
&= O(\varepsilon_n) \qquad (73)
\end{aligned}$$

where we used (65) and (67) for (71), the change of variables $\tilde{z}_j = z_j/\varepsilon_n$ for (72) and Assumption **W.1**.

We define

$$\bar{H}_{\varepsilon_n}(v) = \frac{1}{\varepsilon_n^{p+dk}} \int_\Omega \int_{\substack{\{z_j \,|\, x_0+z_j \notin \Omega\} \\ \text{for any } 1 \leq j \leq k}} \left[\prod_{s=1}^k \eta\left(\frac{|z_s|}{\varepsilon_n}\right)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta\left(\frac{|z_j - z_r|}{\varepsilon_n}\right)\right] |\nabla v(x_0) \cdot z_1|^p \times \rho(x_0)^{k+1} \, \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0.$$

For the latter, we have:

$$\begin{aligned}
\bar{H}_{\varepsilon_n}(v_{\varepsilon_n}) &\leq \frac{C}{\varepsilon_n^{p+dk}} \int_\Omega \int_{\substack{\{z_j \,|\, x_0+z_j \notin \Omega\} \\ \text{for any } 1 \leq j \leq k}} |z_1|^p \cdot \left[\prod_{s=1}^k \eta\left(\frac{|z_s|}{\varepsilon_n}\right)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta\left(\frac{|z_j - z_r|}{\varepsilon_n}\right)\right] \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0 \\
&= C \int_\Omega \int_{\substack{\{\tilde{z}_j \,|\, x_0+\varepsilon_n \tilde{z}_j \notin \Omega\} \\ \text{for any } 1 \leq j \leq k}} |\tilde{z}_1|^p \cdot \left[\prod_{s=1}^k \eta(|\tilde{z}_s|)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta(|\tilde{z}_j - \tilde{z}_r|)\right] \mathrm{d}\tilde{z}_k \cdots \mathrm{d}\tilde{z}_1 \mathrm{d}x_0 \qquad (74) \\
&\leq C \int_\Omega \int_{\substack{\{\tilde{z}_j \,|\, |\tilde{z}_j| \geq \frac{\mathrm{dist}(x_0, \partial\Omega)}{\varepsilon_n}\} \\ \text{for any } 1 \leq j \leq k}} |\tilde{z}_1|^p \cdot \left[\prod_{s=1}^k \eta(|\tilde{z}_s|)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta(|\tilde{z}_j - \tilde{z}_r|)\right] \mathrm{d}\tilde{z}_k \cdots \mathrm{d}\tilde{z}_1 \mathrm{d}x_0
\end{aligned}$$

where we used the change of variables $\tilde{z}_j = z_j/\varepsilon_n$ for (74). Now, by using the dominated convergence and Assumption **W.1**, we get that

$$\bar{H}_{\varepsilon_n}(v_{\varepsilon_n}) = o(1). \qquad (75)$$

We continue by defining:

$$\begin{aligned}
\hat{H}_{\varepsilon_n}(v) &:= \bar{H}_{\varepsilon_n}(v) + \tilde{H}_{\varepsilon_n}(v) \\
&= \frac{1}{\varepsilon_n^{p+dk}} \int_\Omega \int_{(\mathbb{R}^d)^k} \left[\prod_{s=1}^k \eta\left(\frac{|z_s|}{\varepsilon_n}\right)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta\left(\frac{|z_j - z_r|}{\varepsilon_n}\right)\right] |\nabla v(x_0) \cdot z_1|^p \rho(x_0)^{k+1} \, \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0 \\
&= \int_\Omega \int_{(\mathbb{R}^d)^k} \left[\prod_{s=1}^k \eta(|\tilde{z}_s|)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta(|\tilde{z}_j - \tilde{z}_r|)\right] |\nabla v(x_0) \cdot \tilde{z}_1|^p \rho(x_0)^{k+1} \, \mathrm{d}\tilde{z}_k \cdots \mathrm{d}\tilde{z}_1 \mathrm{d}x_0 \qquad (76)
\end{aligned}$$

where we used the change of variables $\tilde{z}_j = z_j/\varepsilon_n$ for (76). We also have

$$\begin{aligned}
\hat{H}_{\varepsilon_n}(v) &= \frac{1}{\varepsilon_n^{p+dk}} \int_\Omega \int_{\mathbb{R}^{dk}} \left[\prod_{s=1}^k \eta\left(\frac{|z_s|}{\varepsilon_n}\right)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta\left(\frac{|z_j - z_r|}{\varepsilon_n}\right)\right] |\nabla v(x_0) \cdot z_1|^p \rho(x_0)^{k+1} \, \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0 \\
&= \int_\Omega \int_{\mathbb{R}^{dk}} \left[\prod_{s=1}^k \eta(|\tilde{z}_s|)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta(|\tilde{z}_j - \tilde{z}_r|)\right] |\nabla v(x_0) \cdot \tilde{z}_1|^p \rho(x_0)^{k+1} \, \mathrm{d}\tilde{z}_k \cdots \mathrm{d}\tilde{z}_1 \mathrm{d}x_0 \\
&= \mathcal{E}_\infty^{(k,p)}(v).
\end{aligned}$$

For $\delta > 0$, we continue by noting that

$$\begin{aligned}
&\left|\hat{H}(v_{\varepsilon_n}) - \mathcal{E}_\infty^{(k,p)}(v^*)\right| \\
&\quad \leq \delta \mathcal{E}_\infty^{(k,p)}(v^*) + C C_\delta \int_\Omega \int_{(\mathbb{R}^d)^k} |z_1|^p \cdot \left[\prod_{s=1}^k \eta(|z_s|)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta(|z_j - z_r|)\right] \\
&\qquad \times |\nabla v_{\varepsilon_n}(x_0) - \nabla v(x_0)|^p \, \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0 \qquad (77) \\
&\quad = \delta \mathcal{E}_\infty^{(k,p)}(v^*) + C C_\delta C(\eta) \int_\Omega |\nabla v_{\varepsilon_n}(x_0) - \nabla v^*(x_0)|^p \, \mathrm{d}x_0
\end{aligned}$$

$$= \delta \mathcal{E}_{\infty}^{(k,p)}(v^*) + C_\delta o(\varepsilon_n) \tag{78}$$

where $C(\eta) = \int_{(\mathbb{R}^d)^k} |z_1|^p \cdot \left[\prod_{s=1}^k \eta(|z_s|)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta(|z_j - z_r|)\right] \mathrm{d}z_k \cdots \mathrm{d}z_1$ which is finite by Assumption **W.1** and where we used (64) for (77) as well as the fact that $\nabla v_{\varepsilon_n} \to \nabla v^*$ in L^p for (78).

We conclude the proof by the following chain of inequalities:

$$|\mathcal{E}_{\varepsilon_n,\mathrm{NL}}^{(k,p)}(v_{\varepsilon_n}) - \mathcal{E}_{\infty}^{(k,p)}(v^*)| \leq T_1 + |H_{\varepsilon_n}(v_{\varepsilon_n}) - \mathcal{E}_{\infty}^{(k,p)}(v^*)|$$

$$\leq CC_\delta \varepsilon_n^p + \delta H_{\varepsilon_n}(v_{\varepsilon_n}) + |H_{\varepsilon_n}(v_{\varepsilon_n}) - \tilde{H}_{\varepsilon_n}(v_{\varepsilon_n})| + |\tilde{H}_{\varepsilon_n}(v_{\varepsilon_n}) - \mathcal{E}_{\infty}^{(k,p)}(v^*)| \tag{79}$$

$$\leq CC_\delta \varepsilon_n^p + \delta H_{\varepsilon_n}(v_{\varepsilon_n}) + C\varepsilon_n + |\hat{H}_{\varepsilon_n}(v_{\varepsilon_n}) - \mathcal{E}_{\infty}^{(k,p)}(v^*)| + |\bar{H}_{\varepsilon_n}(v_{\varepsilon_n})| \tag{80}$$

$$\leq CC_\delta \varepsilon_n^p + \delta H_{\varepsilon_n}(v_{\varepsilon_n}) + C\varepsilon_n + \delta \mathcal{E}_{\infty}^{(k,p)}(v^*) + C_\delta o(\varepsilon_n) + o(1) \tag{81}$$

where we used (70) for (79), (73) for (80) and (75) as well as (78) for (81). By assumption (67), we have that $H_{\varepsilon_n}(v_{\varepsilon_n}) \leq C$ and therefore, by first letting $n \to \infty$ and then $\delta \to 0$, we obtain (68). □

Proposition 4.13 ($\liminf$-inequality for the nonlocal energies)**.** *Assume that **S.1**, **M.1**, **M.2** and **W.1** hold. For every $u \in \mathrm{L}^p(\mu)$ and sequence $u_{\varepsilon_n} \to u$ in $\mathrm{L}^p(\mu)$, we have that:*

$$\liminf_{n\to\infty} \mathcal{E}_{\varepsilon_n,\mathrm{NL}}^{(k,p)}(u_{\varepsilon_n}, \eta) \geq \mathcal{E}_{\infty}^{(k,p)}(v). \tag{82}$$

Proof. In the proof $C > 0$ will denote a constant that can be arbitrarily large, is independent of n, δ and that may change from line to line.

Since (82) is trivial if $\liminf_{n\to\infty} \mathcal{E}_{\varepsilon_n,\mathrm{NL}}^{(k,p)}(u_{\varepsilon_n}, \eta) = \infty$, we might assume without loss of generality (see [94]) that $\sup_{n>0} \mathcal{E}_{\varepsilon_n,\mathrm{NL}}^{(k,p)}(u_{\varepsilon_n}, \eta) \leq C$.

We first assume that ρ is Lipschitz. We will be in the same setting as in [39, Theorem 4.1] and therefore let Ω' be compactly contained in Ω. This implies that there exists $\delta' > 0$ such that $\Omega'' := \bigcup_{x\in\Omega'} B(x,\delta') \subset \Omega$. Furthermore, let J be a positive mollifier supported in $\overline{B(0,1)}$ and for $0 < \delta < \delta'$ as well as $v \in \mathrm{L}^p(\mu)$ we set

$$v_\delta(x) = \int_{\mathbb{R}^d} J_\delta(x - z) v(z) \, \mathrm{d}z.$$

By [61, Theorem C.16], we have that $v_\delta \to v$ in $\mathrm{L}^p(\mu)$, v_δ are smooth and in particular, by Young's convolution inequality, for $\ell \in \{1,2\}$,

$$\|\nabla^\ell v_\delta\|_{\mathrm{L}^\infty(\mathbb{R}^d)} \leq \frac{C}{\delta^{\ell+d}} \|v\|_{\mathrm{L}^1(\Omega)} \leq \frac{C}{\delta^{\ell+d}} \|v\|_{\mathrm{L}^p(\Omega)}. \tag{83}$$

If we therefore set $v = u_{\varepsilon_n}$ and $u_{\varepsilon_n,\delta} := (u_{\varepsilon_n})_\delta$ and insert the latter in (83), we obtain

$$\sup_{\varepsilon_n>0} \sum_{\ell=1}^2 \|\nabla^\ell u_{\varepsilon_n,\delta}\|_{\mathrm{L}^\infty(\mathbb{R}^d)} \leq C \sum_{\ell=1}^2 \frac{1}{\delta^{\ell+d}}$$

where the last inequality follows from the fact that $u_{\varepsilon_n} \to u$ in $\mathrm{L}^p(\Omega)$ implies that $\|u_{\varepsilon_n}\|_{\mathrm{L}^p(\Omega)} \leq C$ uniformly. For fixed $\delta > 0$, we deduce that (67) is satisfied. Furthermore,

$$\int_{\Omega'} |\nabla u_{\varepsilon_n,\delta}(x) - \nabla u_\delta(x)|^p \, \mathrm{d}x = \frac{1}{\delta^{d+1}} \int_{\Omega'} \left| \int_{B(0,\delta)} (\nabla J)\left(\frac{z}{\delta}\right) (u_{\varepsilon_n}(x - z) - u(x - z)) \, \mathrm{d}z \right|^p \mathrm{d}x$$

$$\leq \frac{C}{\delta^{d+1}} \int_{\Omega'} \int_{B(0,\delta)} |u_{\varepsilon_n}(x - z) - u(x - z)|^p \, \mathrm{d}z\mathrm{d}x$$

$$= \frac{C}{\delta^{d+1}} \int_{\Omega'} \int_{B(x,\delta)} |u_{\varepsilon_n}(r) - u(r)|^p \, \mathrm{d}r\mathrm{d}x \tag{84}$$

$$\leq \frac{C}{\delta^{d+1}} \int_{\Omega} |u_{\varepsilon_n}(x) - u(x)|^p \, \mathrm{d}x \tag{85}$$

where we used a change of variables for (84) and the definition of δ' as well as Assumption **S.1** for (85). We conclude from the latter that $\nabla u_{\varepsilon_n,\delta} \to \nabla u_\delta$ in $\mathrm{L}^p(\Omega')$ as $\varepsilon_n \to 0$ and therefore, by Lemma 4.12,

$$\lim_{n\to\infty} \frac{1}{\varepsilon_n^{p+kd}} \int_{(\Omega')^{k+1}} \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\varepsilon_n}\right)\right] |u_{\varepsilon_n,\delta}(x_1) - u_{\varepsilon_n,\delta}(x_0)|^p \prod_{\ell=0}^{k} \rho(x_\ell)\, \mathrm{d}x_k \cdots \mathrm{d}x_0$$

$$= \sigma_\eta \int_{\Omega'} \|\nabla u_\delta(x_0)\|_2^p \, \rho(x_0)^{k+1}\, \mathrm{d}x_0. \tag{86}$$

Let us define

$$a_{\varepsilon_n,\delta} = \frac{1}{\varepsilon_n^{p+dk}} \int_{\mathbb{R}^d} \int_{(\Omega'')^{k+1}} \frac{1}{\delta^d} J\left(\frac{z}{\delta}\right) \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\varepsilon}\right)\right] |u_{\varepsilon_n}(x_1) - u_{\varepsilon_n}(x_0)|^p$$
$$\times \left(\prod_{\ell=0}^{k} \rho(x_\ell) - \prod_{\ell=0}^{k} \rho(x_\ell + z)\right) \mathrm{d}x_k \cdots \mathrm{d}x_0 \mathrm{d}z.$$

We now estimate as follows:

$$\mathcal{E}^{(k,p)}_{\varepsilon_n,\mathrm{NL}}(u_{\varepsilon_n}, \eta) \geq \frac{1}{\varepsilon_n^{p+kd}} \int_{(\Omega'')^{k+1}} \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\varepsilon}\right)\right] |u_{\varepsilon_n}(x_1) - u_{\varepsilon_n}(x_0)|^p \left[\prod_{\ell=0}^{k} \rho(x_\ell)\right] \mathrm{d}x_k \cdots \mathrm{d}x_0$$

$$= \frac{1}{\varepsilon_n^{p+dk}} \int_{\mathbb{R}^d} \int_{(\Omega'')^{k+1}} \frac{1}{\delta^d} J\left(\frac{z}{\delta}\right) \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\varepsilon}\right)\right] |u_{\varepsilon_n}(x_1) - u_{\varepsilon_n}(x_0)|^p$$
$$\times \left[\prod_{\ell=0}^{k} \rho(x_\ell + z)\right] \mathrm{d}x_k \cdots \mathrm{d}x_0 \mathrm{d}z + a_{\varepsilon_n,\delta}$$

$$\geq \frac{1}{\varepsilon_n^{p+dk}} \int_{\mathbb{R}^d} \int_{(\Omega')^{k+1}} \frac{1}{\delta^d} J\left(\frac{z}{\delta}\right) \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|\hat{x}_j - \hat{x}_r|}{\varepsilon}\right)\right] |u_{\varepsilon_n}(\hat{x}_1 - z) - u_{\varepsilon_n}(\hat{x}_0 - z)|^p$$
$$\times \left[\prod_{\ell=0}^{k} \rho(\hat{x}_\ell)\right] \mathrm{d}\hat{x}_k \cdots \mathrm{d}\hat{x}_0 \mathrm{d}z + a_{\varepsilon_n,\delta} \tag{87}$$

$$\geq \frac{1}{\varepsilon_n^{p+dk}} \int_{(\Omega')^{k+1}} \left[\prod_{\ell=0}^{k} \rho(\hat{x}_\ell)\right] \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|\hat{x}_j - \hat{x}_r|}{\varepsilon}\right)\right]$$
$$\times \left|\int_{\mathbb{R}^d} \frac{1}{\delta^d} J\left(\frac{z}{\delta}\right) \left(u_{\varepsilon_n}(\hat{x}_1 - z) - u_{\varepsilon_n}(\hat{x}_0 - z)\right) \mathrm{d}z\right|^p \mathrm{d}\hat{x}_k \cdots \mathrm{d}\hat{x}_0 + a_{\varepsilon_n,\delta} \tag{88}$$

$$= \frac{1}{\varepsilon_n^{p+dk}} \int_{(\Omega')^{k+1}} \left[\prod_{\ell=0}^{k} \rho(\hat{x}_\ell)\right] \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|\hat{x}_j - \hat{x}_r|}{\varepsilon}\right)\right] |u_{\varepsilon_n,\delta}(\hat{x}_1) - u_{\varepsilon_n,\delta}(\hat{x}_0)|^p \, \mathrm{d}\hat{x}_k \cdots \mathrm{d}\hat{x}_0 + a_{\varepsilon_n,\delta} \tag{89}$$

where we used the change of variables $\hat{x}_j = x_j + z$ for $0 \leq j \leq k$ and the fact that, by definition of δ, $\Omega' \subseteq \{w \in \mathbb{R}^d \,|\, w + z \in \Omega'' \text{ for } z \in B(0,\delta)\}$ for (87) as well as Jensen's inequality with probability measure $\nu(A) = \int_A \frac{1}{\delta^d} J\left(\frac{z}{\delta}\right) \mathrm{d}z$ for (88).

Now, using (66), we have

$$|a_{\varepsilon_n,\delta}| \leq \frac{C}{\varepsilon_n^{p+dk}} \int_{\mathbb{R}^d} \int_{(\Omega'')^{k+1}} \frac{1}{\delta^d} J\left(\frac{z}{\delta}\right) \left[\prod_{j=1}^{k}\prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\varepsilon}\right)\right] |u_{\varepsilon_n}(x_1) - u_{\varepsilon_n}(x_0)|^p \, |z| \, \mathrm{d}x_k \cdots \mathrm{d}x_0 \mathrm{d}z$$

(90) $$\leq \frac{C\delta}{\varepsilon_n^{p+dk}} \int_{(\Omega'')^{k+1}} \left[\prod_{j=1}^{k} \prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\varepsilon}\right)\right] |u_{\varepsilon_n}(x_1) - u_{\varepsilon_n}(x_0)|^p \prod_{\ell=0}^{k} \rho(x_\ell)\, \mathrm{d}x_k \cdots \mathrm{d}x_0$$

(91) $$\leq C\delta \mathcal{E}^{(k,p)}_{\varepsilon_n,\mathrm{NL}}(u_{\varepsilon_n}, \eta)$$

where we used Assumption **M.2** for (90). We therefore obtain:

$$\liminf_{n\to\infty} \mathcal{E}^{(k,p)}_{\varepsilon_n,\mathrm{NL}}(u_{\varepsilon_n}, \eta) \geq \liminf_{\delta\to 0} \liminf_{n\to\infty} a_{\varepsilon_n,\delta}$$

(92)

$$+ \liminf_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{\varepsilon_n^{p+dk}} \int_{(\Omega')^{k+1}} \left[\prod_{\ell=0}^{k} \rho(\hat{x}_\ell)\right] \left[\prod_{j=1}^{k} \prod_{r=0}^{j-1} \eta\left(\frac{|\hat{x}_j - \hat{x}_r|}{\varepsilon_n}\right)\right] |u_{\varepsilon_n,\delta}(\hat{x}_1) - u_{\varepsilon_n,\delta}(\hat{x}_0)|^p \,\mathrm{d}\hat{x}_k \cdots \mathrm{d}\hat{x}_0$$

(93)

$$= \liminf_{\delta\to 0} \sigma_\eta \int_{\Omega'} \|\nabla u_\delta(x_0)\|_2^p \, \rho(x_0)^{k+1} \,\mathrm{d}x_0$$

(94)

$$\geq \sigma_\eta \int_{\Omega'} \|\nabla u(x_0)\|_2^p \, \rho(x_0)^{k+1} \,\mathrm{d}x_0$$

where we used (89) for (92), (91) and the fact that the energies $\mathcal{E}^{(k,p)}_{\varepsilon_n,\mathrm{NL}}(u_{\varepsilon_n}, \eta)$ are uniformly bounded as well as (86) for (93) and, since $u_\delta \to u$ in $\mathrm{L}^p(\Omega)$, the fact that $\mathcal{E}^{(k,p)}_\infty$ is lower-semicontinuous for (94). We conclude by noting that Ω' was an arbitrary set compactly contained in Ω so we can take $\Omega' \uparrow \Omega$ in (94) to get (82).

Dealing with the case where ρ is not Lipschitz is done analogously to the proof of [39, Theorem 4.1], i.e. relying on the approximation of continuous functions by a monotone sequence of Lipschitz functions and the monotone convergence theorem to deduce the result. □

Proposition 4.14 (lim sup-inequality for the nonlocal energies)**.** *Assume that* ***S.1****,* ***M.1****,* ***M.2*** *and* ***W.1*** *hold. For every $u \in \mathrm{L}^p(\mu)$, there exists a sequence $\{u_{\varepsilon_n}\}_{n=1}^\infty \subseteq \mathrm{L}^p(\mu)$ such that $u_{\varepsilon_n} \to u$ in $\mathrm{L}^p(\mu)$ and*

(95) $$\limsup_{n\to\infty} \mathcal{E}^{(k,p)}_{\varepsilon_n,\mathrm{NL}}(u_{\varepsilon_n}, \eta) \leq \mathcal{E}^{(k,p)}_\infty(u).$$

In particular, if $u \in \mathrm{C}^\infty(\bar{\Omega})$ then we can choose $u_{\varepsilon_n} = u$.

Proof. In the proof $C > 0$ will denote a constant that can be arbitrarily large, is independent of n, and that may change from line to line.

We start by noting that (95) is trivial if $\mathcal{E}^{(k,p)}_\infty(u) = \infty$ so that we assume $u \in \mathrm{W}^{1,p}(\Omega)$. Furthermore, we are going to apply [39, Remark 2.7], so it is sufficient to verify (95) on a dense subset of $\mathrm{W}^{1,p}(\Omega)$, namely $\mathrm{C}^\infty_\mathrm{c}(\Omega)$.

We first start by assuming that ρ is Lipschitz. Let us define $u_n = u$ and we estimate as follows:

$$\mathcal{E}^{(k,p)}_{\varepsilon_n,\mathrm{NL}}(u, \eta) \leq \frac{1}{\varepsilon_n^p} \int_\Omega \int_{(\mathbb{R}^d)^k} \left[\prod_{s=1}^{k} \eta(|z_s|)\right] \left[\prod_{j=1}^{k} \prod_{r=1}^{j-1} \eta(|z_j - z_r|)\right] |u(x_0 + \varepsilon_n z_1) - u(x_0)|^p$$

(96) $$\times \rho(x_0) \prod_{\ell=1}^{k} \rho(x_0 + \varepsilon_n z_\ell)\, \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0$$

$$\leq \int_\Omega \int_{(\mathbb{R}^d)^k} \left[\prod_{s=1}^{k} \eta(|z_s|)\right] \left[\prod_{j=1}^{k} \prod_{r=1}^{j-1} \eta(|z_j - z_r|)\right] |\nabla u(x_0) \cdot z_1|^p$$

$$\times \left[\rho(x_0) \prod_{\ell=1}^{k} \rho(x_0 + \varepsilon_n z_\ell)\right] \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0$$

$$+ \varepsilon_n \|\rho\|_{\mathrm{L}^\infty}^{k+1} \|u\|_{\mathrm{C}^2}^p \int_\Omega \int_{(\mathbb{R}^d)^k} \left[\prod_{s=1}^k \eta(|z_s|)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta(|z_j - z_r|)\right] \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0$$

$$\leq \int_\Omega \int_{(\mathbb{R}^d)^k} \left[\prod_{s=1}^k \eta(|z_s|)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta(|z_j - z_r|)\right] |\nabla u(x_0) \cdot z_1|^p$$

$$\times \left|\rho(x_0) \prod_{\ell=1}^k \rho(x_0 + \varepsilon_n z_\ell) - \rho(x_0)^{k+1}\right| \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0 + \mathcal{E}_\infty^{(k,p)}(u) + C\varepsilon_n$$

$$\leq C\varepsilon_n \int_\Omega \int_{(\mathbb{R}^d)^k} \left[\prod_{s=1}^k \eta(|z_s|)\right] \left[\prod_{j=1}^k \prod_{r=1}^{j-1} \eta(|z_j - z_r|)\right] |\nabla u(x_0) \cdot z_1|^p \cdot \sum_{\ell=1}^k |z_\ell| \, \mathrm{d}z_k \cdots \mathrm{d}z_1 \mathrm{d}x_0$$

(97) $$+ \mathcal{E}_\infty^{(k,p)}(u) + C\varepsilon_n$$

(98) $$\leq \mathcal{E}_\infty^{(k,p)}(u) + C\varepsilon_n$$

where we used the change of variables $z_j = (x_j - x_0)/\varepsilon_n$ for $1 \leq j \leq k$ for (96), Assumption **W.1**, (65) for (97) and Assumption **W.1** for (98). Taking the limit as $\varepsilon_n \to 0$, we obtain (95).

In order to consider general ρ, we proceed as in [39, Theorem 4.1] which concludes the proof. □

4.3.2 Γ-convergence of the discrete energies

Proposition 4.15 ($\liminf$-inequality in the ill-posed case)**.** *Assume that **S.1, M.1, M.2, W.1, D.1, D.2** and **L.2** hold. Then, $\mathbb{P}$-a.s., for every $(\nu, v) \in \mathrm{TL}^p(\Omega)$ and $\{(\nu_n, v_n)\}_{n=1}^\infty$ with $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^p(\Omega)$, we have*

(99) $$\liminf_{n\to\infty} (\mathcal{SF})_{n,\varepsilon_n}^{(q,p)}((\nu_n, v_n)) \geq (\mathcal{SG})_\infty^{(q,p)}((\nu, v)).$$

Proof. With probability one, we can assume that the conclusions of Theorem 2.3 hold.

Since (99) is trivial if $\liminf_{n\to\infty} (\mathcal{SF})_{n,\varepsilon_n}^{(q,p)}((\nu_n, v_n)) = \infty$, we might assume without loss of generality (see [94]) that $\sup_{n>0} (\mathcal{SF})_{n,\varepsilon_n}^{(q,p)}((\nu_n, v_n)) \leq C$. This implies that $\nu_n = \mu_n$ and, since $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^p(\Omega)$, we have $\nu = \mu$. We start by showing

(100) $$\liminf_{n\to\infty} \mathcal{E}_{n,\varepsilon_n}^{(k,p)}(v_n) \geq \mathcal{E}_\infty^{(k,p)}(v).$$

We follow the three-step decomposition of [39, Theorem 1.1]. First, suppose that $\eta(t) = a$ if $0 \leq t \leq b$ and $\eta(t) = 0$ else where a and b are positive constants. Define $\tilde{\varepsilon}_n = \varepsilon_n - \frac{2\|T_n - \mathrm{Id}\|_{\mathrm{L}^\infty}}{b}$. From [84, Lemma 4.2], we know that $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$ and

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

Using a change of variables and the above, we obtain that

$$\mathcal{E}_{n,\varepsilon_n}^{(k,p)}(v_n) \geq \frac{1}{\varepsilon_n^{p+kd}} \int_{\Omega^{k+1}} \left[\prod_{j=1}^k \prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\tilde{\varepsilon}_n}\right)\right] |v_n \circ T_n(x_1) - v_n \circ T_n(x_0)|^p \left[\prod_{\ell=0}^k \rho(x_\ell)\right] \mathrm{d}x_k \cdots \mathrm{d}x_0$$

$$= \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^{p+kd} \mathcal{E}_{\tilde{\varepsilon}_n,\mathrm{NL}}^{(k,p)}(v_n \circ T_n, \eta).$$

Since $u_n \to u$ in $\mathrm{TL}^p(\Omega)$, we have that $u_n \circ T_n \to u$ in $\mathrm{L}^p(\Omega)$ and we can therefore use Proposition 4.13 to deduce that:

$$\liminf_{n\to\infty} \mathcal{E}_{n,\varepsilon_n}^{(k,p)}(u_n) \geq \liminf_{n\to\infty} \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^{p+kd} \mathcal{E}_{\tilde{\varepsilon}_n,\mathrm{NL}}^{(k,p)}(v_n \circ T_n, \eta) \geq \mathcal{E}_\infty^{(k,p)}(v).$$

Our next step is to assume that $\eta = \sum_{k=1}^\ell \eta_l$ satisfies Assumption **W.1** where η_l are functions of the type considered in the above. Then, as in [39, Theorem 1.1], we use the linearity of the integral to obtain (99).

Our final step is to let η be a general function satisfying Assumption **W.1**. Then, as in [39, Theorem 1.1], we use the monotone convergence theorem and approximation of η by functions as in the previous step to obtain (100).

Now by subadditivity of the $\liminf$ we can conclude (99). $\square$

Proposition 4.16 ($\limsup$-inequality in the well-posed case)**.** *Assume that **S.1**, **M.1**, **M.2**, **W.1**, **D.1**, **D.2** and **L.2** hold. Then, $\mathbb{P}$-a.s., for every $(\nu, v) \in \mathrm{TL}^p(\Omega)$, there exists a sequence $\{(\nu_n, v_n)\}_{n=1}^\infty$ with $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^p(\Omega)$ and*

$$\limsup_{n\to\infty} \mathcal{F}^{(k,p)}_{n,\varepsilon_n}((\nu_n, v_n)) \leq \mathcal{F}^{(k,p)}_{\infty}((\nu, v)). \tag{101}$$

In particular, for $1 \leq k \leq q$ and $v \in \mathrm{C}^\infty(\bar{\Omega})$ with $v(x_i) = y_i$ for $i \leq N$, we can pick $\{(\nu_n, v_n)\}_{n=1}^\infty = \{(\mu_n, v|_{\Omega_n})\}_{n=1}^\infty$.

Proof. With probability one, we can assume that the conclusions of Theorem 2.3 hold.

In the proof $C > 0$ will denote a constant that can be arbitrarily large, is independent of n, and that may change from line to line.

We start by noting that (101) is trivial if $\mathcal{F}^{(k,p)}_\infty(\nu, v) = \infty$ so that we assume $\mathcal{F}^{(k,p)}_\infty(\nu, v) < \infty$ which implies that $\nu = \mu$, $v \in \mathrm{W}^{1,p}(\Omega)$ and $v(x_i) = \ell_i$ for all $i \leq N$. Furthermore, we are going to apply [39, Remark 2.7], so it is sufficient to verify (101) on a dense subset of $\{\mu\} \times \mathrm{W}^{1,p}(\Omega)$, namely we consider $(\mu, v) \in \{\mu\} \times \mathrm{C}^\infty(\bar{\Omega})$ with $v(x_i) = \ell_i$ for all $i \leq N$. We let $\nu_n = \mu_n$ and $v_n = v|_{\Omega_n}$ so $v_n(x_i) = \ell_i$ for all $i \leq N$ and (101) is equivalent to

$$\limsup_{n\to\infty} \mathcal{E}^{(k,p)}_{n,\varepsilon_n}(v_n) \leq \mathcal{E}^{(k,p)}_{\infty}(v).$$

The fact that $(\mu_n, v_n) \to (\mu, v)$ in $\mathrm{TL}^p(\Omega)$ follows analogously from what is shown in [94, Proposition 4.17]: it relies on the fact that v is uniformly continuous as well as $\|T_n - \mathrm{Id}\|_{\mathrm{L}^\infty} \to 0$.

We follow the three-step decomposition of [39, Theorem 1.1]. First, suppose that $\eta(t) = a$ if $0 \leq t \leq b$ and $\eta(t) = 0$ else where a and b are positive constants. Define $\tilde{\varepsilon}_n = \varepsilon_n + \frac{2\|T_n - \mathrm{Id}\|_{\mathrm{L}^\infty}}{b}$. From [39, Theorem 1.1], we know that $\frac{\varepsilon_n}{\tilde{\varepsilon}_n} \to 1$ and similarly to the previous proposition

$$\mathcal{E}^{(k,p)}_{n,\varepsilon_n}(v_n) \leq \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^{p+kd} \mathcal{E}_{\tilde{\varepsilon}_n,\mathrm{NL}}(v_n \circ T_n, \eta). \tag{102}$$

Let $\delta > 0$ and estimate as follows:

$$\begin{aligned}
T_1 &:= |\mathcal{E}^{(k,p)}_{\tilde{\varepsilon}_n,\mathrm{NL}}(v, \eta) - \mathcal{E}^{(k,p)}_{\tilde{\varepsilon}_n,\mathrm{NL}}(v \circ T_n, \eta)| \\
&\leq \frac{1}{\tilde{\varepsilon}_n^{p+kd}} \int_{\Omega^{k+1}} \left[\prod_{j=1}^k \prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\tilde{\varepsilon}_n}\right)\right] \left[\prod_{\ell=0}^k \rho(x_\ell)\right] \\
&\qquad \times ||v(x_1) - v(x_0)|^p - |v_n \circ T_n(x_1) - v_n \circ T_n(x_0)|^p| \, \mathrm{d}x_k \cdots \mathrm{d}x_0 \\
&\leq \frac{C_\delta}{\tilde{\varepsilon}_n^{p+kd}} \int_{\Omega^{k+1}} \left[\prod_{j=1}^k \prod_{r=0}^{j-1} \eta\left(\frac{|x_j - x_r|}{\tilde{\varepsilon}_n}\right)\right] \left[\prod_{\ell=0}^k \rho(x_\ell)\right] \\
&\qquad \times |v(x_1) - v \circ T_n(x_1) - (v(x_0) - v \circ T_n(x_0))|^p \, \mathrm{d}x_k \cdots \mathrm{d}x_0 + \delta \mathcal{E}^{(k,p)}_{\tilde{\varepsilon}_n,\mathrm{NL}}(v, \eta) \qquad (103) \\
&\leq C C_\delta \left(\frac{\|\mathrm{Id} - T_n\|_{\mathrm{L}^\infty}}{\tilde{\varepsilon}_n}\right)^p + \delta \mathcal{E}^{(k,p)}_{\tilde{\varepsilon}_n,\mathrm{NL}}(v, \eta)
\end{aligned}$$

where we used (64) for (103) and the fact that $v \in \mathrm{C}^\infty(\bar{\Omega})$ as well as Assumptions **W.1** and **M.2**. We obtain:

$$\begin{aligned}
\limsup_{n\to\infty} \mathcal{E}^{(k,p)}_{n,\varepsilon_n}(v_n) &\leq \limsup_{n\to\infty} \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^{p+kd} \mathcal{E}^{(k,p)}_{\tilde{\varepsilon}_n,\mathrm{NL}}(v \circ T_n, \eta) &(104)\\
&\leq \limsup_{n\to\infty} \left(\frac{\tilde{\varepsilon}_n}{\varepsilon_n}\right)^{p+kd} \frac{1}{1+\delta} \left(\mathcal{E}^{(k,p)}_{\tilde{\varepsilon}_n,\mathrm{NL}}(v, \eta) + C C_\delta \left(\frac{\|\mathrm{Id} - T_n\|_{\mathrm{L}^\infty}}{\tilde{\varepsilon}_n}\right)^p\right) &(105)
\end{aligned}$$

(106) $$\leq \frac{1}{1+\delta}\mathcal{E}_{\infty}^{(k,p)}(v)$$

where we used (102) for (104), (103) for (105) and the fact that the recovery sequence in Proposition 4.14 for v was v for (106). Letting $\delta \to 0$ proves (101) for η in this form.

Next we proceed as in Proposition 4.15 or [39, Theorem 1.1]: by assuming that $\eta = \sum_{k=1}^{\ell} \eta_l$ satisfies Assumption **W.1** where η_l are functions of the type considered in the above, we use the linearity of the integral to deduce (101); assuming that η is a general function satisfying Assumption **W.1**, we approximate η by functions of the type considered in the previous step and use the monotone convergence theorem to conclude. □

Using the result of Proposition 4.16, we can prove the next straightforward corollary. The key point to note is that, since we have the same recovery sequence for all $1 \leq k \leq q$, we just apply the subadditivity of $\limsup$ to conclude.

Corollary 4.17 ($\limsup$-inequality for the sum of semi-supervised energies in the well-posed case)**.** *Assume that **S.1**, **M.1**, **M.2**, **W.1**, **D.1**, **D.2** and **L.2** hold. Then, $\mathbb{P}$-a.s., for every $(\nu, v) \in \mathrm{TL}^p(\Omega)$, there exists $\{(\nu_n, v_n)\}_{n=1}^{\infty}$ with $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^p(\Omega)$ such that:*

$$\limsup_{n\to\infty} (\mathcal{SF})_{n,\varepsilon_n}^{(k,p)}((\nu_n, v_n)) \leq (\mathcal{SF})_{\infty}^{(k,p)}((\nu, v)).$$

The following proofs use arguments from [84]. For Proposition 4.18, the sum of all energies $\{\mathcal{F}_{n,\varepsilon_n}^{(k,p)}\}_{k=1}^{q}$ has to be considered directly as we plan to use the uniform convergence results for the $k = 1$ case from [84, Lemma 4.5]. In Proposition 4.19, we show that $n\varepsilon_n^p \to \infty$ is the common lower bound for all energies $\{\mathcal{F}_{n,\varepsilon_n}^{(k,p)}\}_{k=1}^{q}$ in order for them to converge to the ill-posed continuum objective functions.

Proposition 4.18 ($\liminf$-inequality for the sum of semi-supervised energies in the well-posed case)**.** *Assume that **S.1**, **M.1**, **M.2**, **W.1**, **D.1**, **D.2** and **L.2** hold. Assume that $n\varepsilon_n^p \to 0$. Then, $\mathbb{P}$-a.s., for every $(\nu, v) \in \mathrm{TL}^p(\Omega)$ and $\{(\nu_n, v_n)\}_{n=1}^{\infty}$ with $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^p(\Omega)$, we have:*

(107) $$\liminf_{n\to\infty} (\mathcal{SF})_{n,\varepsilon_n}^{(q,p)}((\nu_n, v_n)) \geq (\mathcal{SF})_{\infty}^{(q,p)}((\nu, v)).$$

Proof. With probability one, we can assume that the conclusions of Proposition 4.15 and [84, Lemma 4.5] hold.

In the proof $C > 0$ will denote a constant that can be arbitrarily large, is independent of n, and that may change from line to line.

First, by the same argument as in Proposition 4.15, we can assume that $\sup_{n\geq 1} (\mathcal{SF})_{n,\varepsilon_n}^{(k,p)}((\nu_n, v_n)) \leq C$ and therefore $\nu_n = \mu_n$ and $\nu = \mu$. In particular, we also have that $\mathcal{E}_{n,\varepsilon_n}^{(1,p)}(v_n) \leq C$ and, by [84, Lemma 4.5], we deduce the existence of a continuous function $\hat{v}$ such that for any $\Omega' \subset\subset \Omega$, $\max_{\{i\leq n_k \mid x_i \in \Omega'\}} |v_{n_k}(x_i) - \hat{v}(x_i)| \to 0$: this implies that $\hat{v}(x_i) = \ell_i$ for all $i \leq N$ with probability one. We also note that $v = \hat{v}$ (in particular, $v(x_i) = \ell_i$ for all $i \leq N$) and (107) reduces to proving

$$\liminf_{n\to\infty} \sum_{k=1}^{q} \lambda_k \mathcal{E}_{n,\varepsilon_n}^{(k,p)}(v_n) \geq \sum_{k=1}^{q} \lambda_k \mathcal{E}_{\infty}^{(k,p)}(v) = \sum_{k=1}^{q} \lambda_k \mathcal{G}_{\infty}^{(k,p)}((\mu, v)).$$

By Proposition 4.15, we know that $\liminf_{n\to\infty} \mathcal{E}_{n,\varepsilon_n}^{(k,p)}(v_n) \geq \mathcal{G}_{\infty}((\mu, v))$ so that:

$$\liminf_{n\to\infty} \sum_{k=1}^{q} \lambda_k \mathcal{E}_{n,\varepsilon_n}^{(k,p)}(v_n) \geq \sum_{k=1}^{q} \lambda_k \liminf_{n\to\infty} \mathcal{E}_{n,\varepsilon_n}^{(k,p)}(v_n)$$
$$\geq \sum_{k=1}^{q} \lambda_k \mathcal{G}_{\infty}^{(k,p)}((\mu, v)).$$
□

Proposition 4.19 (lim sup-inequality in the ill-posed case)**.** *Assume that **S.1**, **M.1**, **M.2**, **W.1**, **D.1**, **D.2** and **L.2** hold. Assume that $n\varepsilon_n^p \to \infty$. Then, $\mathbb{P}$-a.s., for every $(\nu, v) \in \mathrm{TL}^p(\Omega)$, there exists $\{(\nu_n, v_n)\}_{n=1}^\infty$ with $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^p(\Omega)$ such that:*

$$\limsup_{n\to\infty} \mathcal{F}^{(k,p)}_{n,\varepsilon_n}((\nu_n, v_n)) \leq \mathcal{G}^{(k,p)}_\infty((\nu, v)). \tag{108}$$

In particular, for $1 \leq k \leq q$ and $v \in \mathrm{C}^\infty(\bar{\Omega})$, we can pick $\{(\nu_n, v_n)\}_{n=1}^\infty = \{(\mu_n, \hat{v}_n)\}_{n=1}^\infty$ where $\hat{v}_n(x_i) = y_i$ for $i \leq N$ and $\hat{v}_n = v|_{\Omega_n}$ else.

Proof. With probability one, we can assume that the conclusions of Proposition 4.16 hold.

In the proof $C > 0$ will denote a constant that can be arbitrarily large, is independent of n, and that may change from line to line.

We start by noting that (108) is trivial if $\mathcal{G}^{(k,p)}_\infty((\nu, v)) = \infty$ so that we assume $\mathcal{G}^{(k,p)}_\infty((\nu, v)) < \infty$ which implies that $\nu = \mu$ and $v \in \mathrm{W}^{1,p}(\Omega)$. Furthermore, we are going to apply [39, Remark 2.7], so it is sufficient to verify (108) on a dense subset of $\{\mu\} \times \mathrm{W}^{1,p}(\Omega)$, namely we consider $(\mu, v) \in \{\mu\} \times \mathrm{C}^\infty(\bar{\Omega})$. We let $\nu_n = \mu_n$ and $\hat{v}_n = v|_{\Omega_n}$.

By repeating the proof of Proposition 4.16, we can show that $(\nu_n, \hat{v}_n) \to (\nu, v)$ in $\mathrm{TL}^p(\Omega)$ and

$$\limsup_{n\to\infty} \mathcal{E}^{(k,p)}_{n,\varepsilon_n}(\hat{v}_n) \leq \mathcal{G}^{(k,p)}_\infty((\mu, v)). \tag{109}$$

The subtlety of (109) compared to (101) is that $\hat{v}_n$ does not necessarily satisfy $\hat{v}_n(x_i) = \ell_i$ for all $i \leq N$ since this condition is not imposed on v. We note that $\|\hat{v}_n\|_{\mathrm{L}^\infty} \leq C$ since $\hat{v}_n = v|_{\Omega_n}$ and $v \in \mathrm{C}^\infty(\bar{\Omega})$.

Define (μ_n, v_n) with

$$v_n(x_i) = \begin{cases} y_i & \text{if } i \leq N, \\ \hat{v}_n(x_i) & \text{else.} \end{cases}$$

Again, we have $\|v_n\|_{\mathrm{L}^\infty} \leq C$ and, using the arguments of [94, Proposition 4.24], we can show that $(\mu_n, v_n) \to (\mu, v)$ in $\mathrm{TL}^p(\Omega)$. Since $v_n(x_i) = \ell_i$ for all $i \leq N$, in order to show (108), it therefore is sufficient to show that

$$\lim_{n\to\infty} \Bigg(\underbrace{\mathcal{F}^{(k,p)}_{n,\varepsilon_n}(\mu_n, v_n) - \mathcal{E}^{(k,p)}_{n,\varepsilon_n}(\hat{v}_n)}_{=:T_2} \Bigg) = 0.$$

We estimate as follows:

$$|T_2| \leq \frac{1}{n^{k+1}\varepsilon_n^{p+kd}} \sum_{i_0,\cdots,i_k=1}^n \left[\prod_{j=1}^k \prod_{r=0}^{j-1} \eta\left(\frac{|x_{i_j} - x_{i_r}|}{\varepsilon_n} \right) \right] ||v_n(x_{i_1}) - v_n(x_{i_0})|^p - |\hat{v}_n(x_{i_1}) - \hat{v}_n(x_{i_0})|^p|$$

By definition of v_n, for $(i_0, \cdots, i_k) \in S := \{(i_0, \cdots, i_k) \,|\, N \leq i_j \leq n \text{ for all } 0 \leq j \leq 1\}$, the corresponding term in the above sum vanishes. This means that we need to consider all indices in

$$S^c = \{(i_0, \cdots, i_k) \,|\, \text{there exists } 0 \leq j \leq 1 \text{ such that } 1 \leq i_j \leq N\}.$$

Summing over all sets in S^c therefore yields:

$$\begin{aligned} T_2 &\leq \frac{1}{n^{k+1}\varepsilon_n^{p+kd}} \sum_{t=0}^1 \sum_{i_t=1}^N \sum_{\substack{i_s=1 \\ s\neq t}}^n \left[\prod_{j=1}^k \prod_{r=0}^{j-1} \eta\left(\frac{|x_{i_j} - x_{i_r}|}{\varepsilon_n} \right) \right] ||v_n(x_{i_1}) - v_n(x_{i_0})|^p - |\hat{v}_n(x_{i_1}) - \hat{v}_n(x_{i_0})|^p| \\ &\leq \frac{C}{n\varepsilon_n^p} \sum_{t=0}^1 \sum_{i_t=1}^N \frac{1}{n^k \varepsilon_n^{dk}} \sum_{\substack{i_s=1 \\ s\neq t}}^n \left[\prod_{j=1}^k \prod_{r=0}^{j-1} \eta\left(\frac{|x_{i_j} - x_{i_r}|}{\varepsilon_n} \right) \right] \qquad (110) \\ &\leq \frac{C}{n\varepsilon_n^p} \sum_{t=0}^1 \sum_{i_t=1}^N \frac{1}{n^k \varepsilon_n^{dk}} \sum_{\substack{i_s=1 \\ s\neq t}}^n \left[\prod_{j=1}^k \eta\left(\frac{|x_{i_j} - x_{i_{j-1}}|}{\varepsilon_n} \right) \right]. \end{aligned}$$

For $t \in \{0,1\}$, using $\eta(s) = 0$ for all $|s| > 1$,

$$\frac{1}{n^k \varepsilon_n^{dk}} \sum_{\substack{i_s=1\\ s\neq t}}^n \left[\prod_{j=1}^k \eta\left(\frac{|x_{i_j} - x_{i_{j-1}}|}{\varepsilon_n}\right)\right]$$
$$\leq \frac{\eta(0)^k}{n^k \varepsilon_n^{dk}} \#\{(i_0, \cdots, i_{t-1}, i_{t+1}, \cdots, i_k) \,|\, |x_{i_j} - x_{i_{j-1}}| < \varepsilon_n \text{ for } 0 \leq j \leq k\}.$$

Now, for an element in $(i_0, \cdots, i_k) \in \{(i_0, \cdots, i_{t-1}, i_{t+1}, \cdots, i_k) \,|\, |x_{i_j} - x_{i_{j-1}}| < \varepsilon_n \text{ for } 0 \leq j \leq k\} =: \hat{S}$, we have $x_{i_{t-1}} \in B(x_{i_t}, \varepsilon_n)$, $x_{i_{t-2}} \in B(x_{i_{t-1}}, \varepsilon_n)$ until $x_{i_0} \in B(x_{i_1}, \varepsilon_n)$ as well as $x_{i_{t+1}} \in B(x_{i_t}, \varepsilon_n)$, $x_{i_{t+2}} \in B(x_{i_{t+1}}, \varepsilon_n)$ until $x_{i_k} \in B(x_{i_{k-1}}, \varepsilon_n)$. Hence $x_{i_j} \in B(x_{i_t}, k\varepsilon_n)$ for all j. This shows that

$$\begin{aligned}\#\hat{S} &\leq \sum_{z_1, \cdots, z_k \in \Omega_n} \prod_{j=1}^k \mathbb{1}_{B(x_{i_t}, bk\varepsilon_n)}(z_j) \\ &= \left(n\mu_n(B(x_{i_t}, k\varepsilon_n))\right)^k.\end{aligned}$$

Using the latter, we continue estimating:

$$\begin{aligned}\frac{1}{n^k \varepsilon_n^{dk}} \sum_{\substack{i_s=1\\ s\neq t}}^n \prod_{j=1}^k \eta\left(\frac{|x_{i_j} - x_{i_{j-1}}|}{\varepsilon_n}\right) &\leq \frac{\eta(0)^k}{n^k \varepsilon_n^{dk}} \#\hat{S} \\ &\leq C\left(\varepsilon_n^{-d} \mu_n(x_{i_t}, k\varepsilon_n)\right)^k \\ &= C\left(\varepsilon_n^{-d} \int_\Omega \mathbb{1}_{\{|T_n(x) - x_{i_t}| < k\varepsilon_n\}} \rho(x)\,\mathrm{d}x\right)^k \\ &\leq C\left(\varepsilon_n^{-d} \int_\Omega \mathbb{1}_{\{|x - x_{i_t}| < k\varepsilon_n - \|T_n - \mathrm{Id}\|_{\mathrm{L}^\infty}\}} \rho(x)\,\mathrm{d}x\right)^k \\ &\leq C\left(\mathrm{Vol}(B(0,1)) \left(\frac{k\varepsilon_n - \|T_n - \mathrm{Id}\|_{\mathrm{L}^\infty}}{\varepsilon_n}\right)^d\right)^k \qquad (111) \\ &\leq C \qquad (112)\end{aligned}$$

where we used Assumption **M.2** for (111) and Assumption **L.2** for (112).

Inserting (112) in (110), we obtain that

$$T_2 \leq \frac{C}{n\varepsilon_n^p}$$

from which we deduce that $T_2 \to 0$ and (108). □

The next corollary is the analogue to Corollary 4.17 and is proved in the same manner.

Corollary 4.20 (lim sup-inequality for the sum of semi-supervised energies in the ill-posed case)**.** *Assume that* ***S.1**, **M.1**, **M.2**, **W.1**, **D.1**, **D.2** and **L.2** hold. Assume that $n\varepsilon_n^p \to \infty$. Then, $\mathbb{P}$-a.s., for every $(\nu, v) \in \mathrm{TL}^p(\Omega)$, there exists $\{(\nu_n, v_n)\}_{n=1}^\infty$ with $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^p(\Omega)$ such that:*

$$\limsup_{n\to\infty} (\mathcal{SF})_{n,\varepsilon_n}^{(k,p)}((\nu_n, v_n)) \leq (\mathcal{SG})_\infty^{(k,p)}((\nu, v)).$$

We conclude with a lemma summarizing our Γ-convergence results for our semi-supervised objectives.

Lemma 4.21 (Γ-convergence of energies)**.** *Assume that* ***S.1**, **M.1**, **M.2**, **W.1**, **D.1**, **D.2** and **L.2** hold. If $n\varepsilon_n^p \to 0$, then, $\mathbb{P}$-a.e, $(\mathcal{SF})_{n,\varepsilon_n}^{(k,p)}$ Γ-converges to $(\mathcal{SF})_\infty^{k,p}$ in $\mathrm{TL}^p(\Omega)$. If $n\varepsilon_n^p \to \infty$, then, $\mathbb{P}$-a.s., $(\mathcal{SF})_{n,\varepsilon_n}^{(k,p)}$ Γ-converges to $(\mathcal{SG})_\infty^{k,p}$ in $\mathrm{TL}^p(\Omega)$.*

4.3.3 Compactness and proof of Theorem 3.3

The following lemma is inspired by [94].

Lemma 4.22 (Uniform bound of energies of minimizers)**.** *Assume that* ***S.1, M.1, M.2, W.1, D.1, D.2*** *and* ***L.2*** *hold. Let (μ_n, u_n) be minimizers of $(\mathcal{SF})^{(q,p)}_{n,\varepsilon_n}$. Then, $\mathbb{P}$-a.s., there exists $C > 0$ such that*

$$\sup_{n>0} (\mathcal{SF})^{(q,p)}_{n,\varepsilon_n}((\mu_n, u_n)) \leq C.$$

Proof. With probability one, we can assume that the conclusions of Lemma 4.21 hold.

In the proof $C > 0$ will denote a constant that can be arbitrarily large, is independent of n, and that may change from line to line. We are going to follow the proof of [94, Lemma 4.25]

Let $v \in \mathrm{C}^\infty_\mathrm{c}(\Omega)$ be a function that interpolates the points $\{(x_i, \ell_i)\}_{i=1}^\infty$. Then, $v \in \mathrm{W}^{1,p}(\Omega)$ so in particular, by Assumption **M.2**, there exists C_0 such that

$$(\mathcal{SF})^{(q,p)}_\infty((\mu, v)) < C_0.$$

By Lemma 4.21, we can pick a recovery sequence $\{v_n\}_{n=1}^\infty$ for v such that:

$$\begin{aligned} \lim_{n\to\infty} h_n &:= \lim_{n\to\infty} \sup_{m\geq n} (\mathcal{SF})^{(q,p)}_{m,\varepsilon_m}((\mu_m, v_m)) \\ &= \limsup_{n\to\infty} (\mathcal{SF})^{(q,p)}_{n,\varepsilon_n}((\mu_n, v_n)) \\ &\leq (\mathcal{SF})^{(q,p)}_\infty((\mu, v)) \\ &< C_0 \end{aligned}$$

(since $(\mathcal{SF})^{(q,p)}_\infty((\mu, v)) = (\mathcal{SG})^{(q,p)}_\infty((\mu, v))$). Let $h := \limsup_{n\to\infty} (\mathcal{SF})^{(q,p)}_{n,\varepsilon_n}((\mu_n, v_n))$ and let $\bar{\varepsilon} = C_0 - h > 0$. Then, there exists n_0 such that for all $n \geq n_0$, $h_n - h < \bar{\varepsilon}/2$, which means that

$$h_n = \sup_{m\geq n} (\mathcal{SF})^{(q,p)}_{m,\varepsilon_m}(v_m) < h + \bar{\varepsilon}/2 < C_0.$$

Using the latter, we have

$$\begin{aligned} \sup_{n>0} (\mathcal{SF})^{(q,p)}_{n,\varepsilon_n}((\mu_n, v_n)) &= \max\left\{ (\mathcal{SF})^{(q,p)}_{1,\varepsilon_1}((\mu_1, v_1)), \dots, (\mathcal{SF})^{(q,p)}_{n_0,\varepsilon_{n_0}}((\mu_{n_0}, v_{n_0})), \sup_{n\geq n_0} (\mathcal{SF})^{(q,p)}_{n,\varepsilon_n}((\mu_n, v_n)) \right\} \\ &\leq C. \end{aligned}$$

Since $\{u_n\}_{n=1}^\infty$ are minimizers, we use $(\mathcal{SF})^{(q,p)}_{n,\varepsilon_n}((\mu_n, u_n)) \leq (\mathcal{SF})^{(q,p)}_{n,\varepsilon_n}((\mu_n, v_n))$ to conclude. □

Proof of Theorem 3.3. With probability one, we can assume that the conclusions of Lemmas 4.21 and 4.22 hold.

In the proof $C > 0$ will denote a constant that can be arbitrarily large, is independent of n, and that may change from line to line.

By Lemma 4.22, we know that there exists C such that $\sup_{n>0} (\mathcal{SF})^{(k,p)}_{n,\varepsilon_n}((\mu_n, u_n)) < C$ and in particular, $\mathcal{E}^{(1,p)}_{n,\varepsilon_n}(u_n)$ is uniformly bounded. Furthermore, analogously to what is described in the proof of [84, Theorem 2.1], we know that $\sup_{n>0} \|u_n\|_{\mathrm{L}^\infty} < C$ with probability 1. We can therefore apply [84, Proposition 4.4] to obtain a subsequence $\{n_r\}_{r=1}^\infty$ and $(\mu, u) \in \mathrm{TL}^p(\Omega)$ such that $(\mu_{n_r}, u_{n_r}) \to (\mu, u)$ in $\mathrm{TL}^p(\Omega)$.

1. Since $n\varepsilon_n^p \to 0$, by [84, Lemma 4.5], we know that u is continuous and, for every $\Omega' \subset\subset \Omega$, we have that $\max_{\{s\leq n_r \,|\, x_s\in\Omega'\}} |u(x_s) - u_{n_r}(x_s)| \to 0$ and, with probability 1, $u(x_i) = \ell_i$ for all $i \leq N$. By Lemma 4.21 and Proposition 2.6, we also have that (μ, u) is a minimizer of $(\mathcal{SF})^{(q,p)}_\infty$. Finally, by the uniqueness of the minimizer of $(\mathcal{SF})^{(q,p)}_\infty$, we conclude that the whole sequence (μ_n, u_n) converges to (μ, n) in $\mathrm{TL}^p(\Omega)$ and for every $\Omega' \subset\subset \Omega$, we have that $\max_{\{s\leq n \,|\, x_s\in\Omega'\}} |u(x_s) - u_n(x_s)| \to 0$.

2. By Lemma 4.21, Proposition 2.6 and the assumption $n\varepsilon_n^p \to \infty$, (μ, u) is a minimizer of $(\mathcal{SG})^{(k,p)}_\infty$. □

4.4 Higher-order hypergraph learning

The proofs in this section are simple corollaries from the results in [94]. In contrast to the discrete-continuum nonlocal-continuum local decomposition used for the proofs in Section 4, everything in this section relies on spectral convergence results between the discrete Laplace operators Δ_{n,ε_n} and its continuum counterpart Δ_ρ.

For our first result, the proof follows from an application of [94, Proposition 4.21]. The key observation is that for any ε_n, $p > 0$ and $v \in \mathrm{C}^\infty(\bar{\Omega})$ with $v(x_i) = y_i$ for $i \leq N$, we can pick $\{(\nu_n, v_n)\}_{n=1}^\infty = \{(\mu_n, v|_{\Omega_n})\}_{n=1}^\infty$ and hence the same recovery sequence for $\mathcal{J}^{(p_k)}_{n,\Delta_{n,\varepsilon_n^{(k)}}}$ with $1 \leq k \leq q$. This allows us to use the subadditivity of $\limsup$ to deduce the result.

Proposition 4.23 ($\limsup$-inequality for the sum of semi-supervised energies in the well-posed case)**.** *Assume that **S.2**, **M.1**, **M.2**, **W.1**, **D.1** and **D.2** hold. Let $q \geq 1$, $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$ with $p_1 \leq \cdots \leq p_q$ and $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$ with $\varepsilon_n^{(1)} > \cdots > \varepsilon_n^{(q)}$. Assume that $\varepsilon_n^{(q)}$ satisfies **L.3** and that $\rho \in \mathrm{C}^\infty$. Then, $\mathbb{P}$-a.s., for every $(\nu, v) \in \mathrm{TL}^2(\Omega)$, there exists a sequence $\{(\nu_n, v_n)\}_{n=1}^\infty$ with $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^2(\Omega)$ and*

$$\limsup_{n\to\infty} (\mathcal{SJ})^{(q,P)}_{n,E_n}((\nu_n, v_n)) \leq (\mathcal{SJ})^{(q,P)}_\infty((\nu, v)).$$

The next result is shown analogously to Proposition 4.18. In particular, one relies on the compactness result [94, Proposition 4.13]: we only require that our smallest length-scale $\varepsilon_n^{(q)}$ satisfies the appropriate upper bound and that the its associated power p_q scales correctly with the dimension of Ω. Then, the problem reduces to using the superadditivity of $\liminf$ and [94, Theorem 4.14].

Proposition 4.24 ($\liminf$-inequality for the sum of semi-supervised energies in the well-posed case)**.** *Assume that **S.2**, **M.1**, **M.2**, **W.1**, **D.1** and **D.2** hold. Let $q \geq 1$, $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$ with $p_1 \leq \cdots \leq p_q$ and $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$ with $\varepsilon_n^{(1)} > \cdots > \varepsilon_n^{(q)}$. Assume that $\varepsilon_n^{(q)}$ satisfies **L.3**, that $n \cdot (\varepsilon_n^{(q)})^{p_q/2-1/2}$ is bounded and that $p_q > \frac{5}{2}d + 4$. Then, $\mathbb{P}$-a.s., for every sequence $\{(\nu_n, v_n)\}_{n=1}^\infty \subseteq \mathrm{TL}^2(\Omega)$ with $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^2(\Omega)$, we have*

$$\liminf_{n\to\infty} (\mathcal{SJ})^{(q,P)}_{n,E_n}((\nu_n, v_n)) \geq (\mathcal{SJ})^{(q,P)}_\infty((\nu, v)).$$

For the next result, we again rely on the fact that [94, Proposition 4.24] implies that the same recovery can be chosen for all $\mathcal{J}^{(p_k)}_{n,\Delta_{n,\varepsilon_n^{(k)}}}$ with $1 \leq k \leq q$. In particular, we need to assume that all $\varepsilon_n^{(k)}$ satisfy an appropriate lower bound and that their associated powers p_k scale correctly with the dimension of Ω. We then conclude using the subadditivity of $\limsup$.

Proposition 4.25 ($\limsup$-inequality for the sum of semi-supervised energies in the ill-posed case)**.** *Assume that **S.2**, **M.1**, **M.2**, **W.1**, **D.1** and **D.2** hold. Let $q \geq 1$, $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$ with $p_1 \leq \cdots \leq p_q$ and $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$ with $\varepsilon_n^{(1)} > \cdots > \varepsilon_n^{(q)}$. Assume that $\rho \in \mathrm{C}^\infty$ and that for $\varepsilon_n^{(q)}$ satisfies **L.2** as well as $n(\varepsilon_n^{(q)})^{2p_q} \to \infty$. Then, $\mathbb{P}$-a.s., for every $(\nu, v) \in \mathrm{TL}^2(\Omega)$, there exists a sequence $\{(\nu_n, v_n)\}_{n=1}^\infty$ with $(\nu_n, v_n) \to (\nu, v)$ in $\mathrm{TL}^2(\Omega)$ and*

$$\limsup_{n\to\infty} (\mathcal{SJ})^{(q,P)}_{n,E_n}((\nu_n, v_n)) \leq (\mathcal{SK})^{(q,P)}_\infty((\nu, v)).$$

Summarizing all our previous results and using the subadditivity of $\liminf$ in conjunction with [94, Proposition 4.22], we obtain the following result.

Lemma 4.26 (Γ-convergence of energies)**.** *Assume that **S.2**, **M.1**, **M.2**, **W.1**, **D.1**, **D.2** hold. Let $q \geq 1$, $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$ with $p_1 \leq \cdots \leq p_q$ and $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$ with $\varepsilon_n^{(1)} > \cdots > \varepsilon_n^{(q)}$. Assume that $\rho \in \mathrm{C}^\infty$.*

1. *Assume that $\varepsilon_n^{(q)}$ satisfies **L.3**, that $n \cdot (\varepsilon_n^{(q)})^{p_q/2-1/2}$ is bounded and that $p_q > \frac{5}{2}d + 4$. Then, $\mathbb{P}$-a.s., $(\mathcal{SJ})^{(q,P)}_{n,E_n}$ Γ-converges to $(\mathcal{SJ})^{(q,P)}_\infty$.*

2. *Assume that $\varepsilon_n^{(q)}$ satisfies **L.2** as well as $n(\varepsilon_n^{(q)})^{2p_q} \to \infty$. Then, $\mathbb{P}$-a.s., $(\mathcal{SJ})^{(q,P)}_{n,E_n}$ Γ-converges to $(\mathcal{SK})^{(q,P)}_\infty$.*

Proof of Theorem 3.4. The proof is analogous to the proof of Theorem 3.3.

In particular, for the well-posed case, we use [94, Proposition 4.17 and Lemma 4.25] to obtain a uniform bound on the L^2-norms of u_n. Then, uniform and TL^2-convergence of a subsequence of u_n to some continuous u follows from [94, Proposition 4.13 and Theorem 4.14]. By the uniqueness of the minimizer, Lemma 4.26 and Proposition 2.6, the result follows.

For the ill-posed case, convergence of a subsequence in TL^2 to some u follows from [94, Theorem 4.14]. Again, Lemma 4.26 and Proposition 2.6 allow us to conclude. $\square$

5 Numerical Experiments

Multiscale Laplace learning has demonstrated strong empirical performance on point cloud data, outperforming many existing graph-based semi-supervised learning methods [65]. Building on this, and given that we approximate HOHL using this framework, our evaluation focuses on sensitivity analyses. In particular, we show that choosing exponents $p_\ell = \ell$ leads to improved performance over constant-exponent settings, highlighting the benefit of applying higher-order regularization at finer scales.

In particular, we include the following experiments on four datasets of various sizes and difficulty: iris [31], digits [4], Salinas A [1], MNIST [59]. We summarize all the notation used in Table 2.

q-Experiment. For $1 \leq q \leq 5$, we test (6) with various configurations of weight and power coefficients. In particular, first, we pick $\varepsilon^{(1)} \geq \varepsilon^{(2)} \geq \varepsilon^{(3)} \geq \varepsilon^{(4)} \geq \varepsilon^{(5)}$ and build the Laplacian matrices $\Delta_{n,\varepsilon^{(\ell)}}$ for $1 \leq \ell \leq 5$. Then, we setup different models with constant coefficients $\lambda_\ell = 1$ (CC), slowly increasing coefficients $\lambda_\ell = \ell$ (SC) or quickly increasing coefficients $\lambda_\ell = \ell^2$ (QC) as well as constant powers $p_\ell = 1$ (CP) or increasing powers $p_\ell = \ell$ for $1 \leq \ell \leq 4$ (IP). The aim of this experiment is to analyze the performance of our model as a function of q.

j-Experiment. For $1 \leq q \leq 3$, we pick $\varepsilon_n^{(1)} \geq \varepsilon_n^{(2)} \geq \varepsilon_n^{(3)}$ and build the Laplacian matrices $\Delta_{n,\varepsilon^{(\ell)}}$ for $1 \leq \ell \leq 3$. Then, we set up different models for $1 \leq j \leq 4$ with $\lambda_1 = 1$, $\lambda_2 = j^2$, $\lambda_3 = (j+1)^2$ (VQC(q) – we let $q \in \{2, 3\}$) as well as $p_1 = 1$, $p_2 = 2$ and $p_3 = 3$. The aim of this experiment is to compare the performance of our model as a function of the coefficients λ_ℓ for a fixed q.

Term / Abbreviation	**q-Experiment**	**j-Experiment**
Aim of experiment	Analysis of HOHL as a function of maximum powers q	Analysis of HOHL as a function of coefficients λ_ℓ
ℓ	Index over scales $1 \leq \ell \leq q$	Same meaning
q	Number of Laplacians $1 \leq q \leq 5$	Number of Laplacians $2 \leq q \leq 3$
j	—	Coefficients λ_ℓ are a function of parameter $1 \leq j \leq 4$
$\varepsilon^{(\ell)}$	Scale for ℓ-th ε-graph Laplacian	Same meaning
$k^{(\ell)}$	Scale for ℓ-th kNN-graph Laplacian	Same meaning
$\Delta_{n,\varepsilon^{(\ell)}}$	ℓ-th ε-graph Laplacian	Same meaning
λ_ℓ	Fixed or increasing (1 or ℓ or ℓ^2)	Varies with j: $\lambda_1 = 1, \lambda_2 = j^2, \lambda_3 = (j+1)^2$
p_ℓ	Constant or increasing (1 or ℓ)	Increasing: $p_\ell = \ell$
CC	$\lambda_\ell = 1$	—
SC	$\lambda_\ell = \ell$	—
QC	$\lambda_\ell = \ell^2$	—
CP	$p_\ell = 1$	—
IP	$p_\ell = \ell$	$p_l = \ell$
VQC(q)	—	For $q = 2$: $\lambda_1 = 1, \lambda_2 = j^2$. For $q = 3$: $\lambda_1 = 1, \lambda_2 = j^2$, $\lambda_3 = (j+1)^2$.

Table 2: Terminology used in the q- and j-experiments.

We always use the full dataset as nodes in the graph construction. For MNIST, we use the same data embedding as in [17]. For the smaller datasets, iris and digits, we use ε-graphs with weights $w_{\varepsilon^{(\ell)},ij} = \exp\left(\frac{-4|x_i-x_j|^2}{\left(\varepsilon^{(\ell)}\right)^2}\right)$. To illustrate that our model works with different (hyper)graph types and in order to speed-up computations, we rely on k-nearest neighbors (kNN) graphs for the large datasets (naturally substituting the sequence $\varepsilon^{(1)} \geq \varepsilon^{(2)} \geq \varepsilon^{(3)} \geq \varepsilon^{(4)} \geq \varepsilon^{(5)}$ with $k^{(1)} \geq k^{(2)} \geq k^{(3)} \geq k^{(4)} \geq k^{(5)}$) with weights $w_{k^{(\ell)},ij} = \exp\left(\frac{-4|x_i-x_j|^2}{d_{k^{(\ell)}}(x_i)^2}\right)$ where $d_{k^{(\ell)}}(x_i)$ denotes the distance from x_i to its $k^{(\ell)}$-th nearest neighbor. Each experiment is conducted over 100 trials and we report the mean accuracy and standard deviation (in brackets) of our results in percentages. In particular, for each trial we re-sample labelled points to be used as fixed constraints in the learning (the same constraints as in (2)). We vary the labelling rate from 0.02 to 0.8 (for the Salinas A dataset, the rate parameter goes from 1 to 100 and denotes the number of labelled points per class).

Since we show that hypergraph learning problem and HOHL behave asymptotically like graph problems in Theorems 3.2, 3.3 and 3.4 (see Figure 2), it is relevant to consider other graph algorithms for fair comparisons: Laplace learning [103], Poisson learning [17], Fractional Laplace (FL) learning [94] with $s = 2$ and $s = 3$ (only on iris), Weighted Nonlocal Laplacian (WNLL) [79], p-Laplace learning [32], Random Walk (RW) [101], Centered Kernel (CK) [63], Sparse LP (SLP) [56] and Properly Weighted Graph-Laplacian [19]. Lastly, we note that no methodical hyperparameter optimization has been performed for the choice of $\varepsilon^{(\ell)}$ and $k^{(\ell)}$. The result highlights are displayed in Tables 3, 4, 5, 6, 7 and 8. The complete results can be found in Appendix 7.

We note that our proposed models mostly outperform the other graph SSL models, especially when the labelling rate, q and j are large enough (see also Tables 10, 12, 14 and 16). This signifies that the multiscale/hypergraph structure can be efficiently leveraged when solving a discrete learning problem on point clouds. Furthermore, our method is robust with respect to the choice of parameters p_ℓ and λ_ℓ.

The model configurations with increasing powers (IP) outperform all other versions of our model (5) (see also Tables 9, 11, 13 and 15). This confirms the idea that the choice of $p_\ell = \ell$ is more efficient than $p_\ell = 1$. In particular, this means that our proposed HOHL model (5)/(6) is more effective than the hypergraph learning model (2) because of higher-regularization.

Finally, we note that a choice of $q = 2, 3$ is often sufficient to have a significant increase in performance (see Tables 9, 11 and 13) which is analogous to observations in [65]. This is an important finding as higher values of q imply a notable increase in computation time: for each additional $\varepsilon^{(\ell)}$ or $k^{(\ell)}$, one has to compute a new set of hyperedges/a new Laplacian as well as an increasingly costly matrix product. A similar statement can be made for j (see Tables 10 and 12): increasing the weights λ_ℓ is effective up to a certain point at which the performance of our models does not improve. For this reason, we only consider a restricted set of q and j when dealing with larger datasets.

RATE	LAPLACE	POISSON	**IP-QC**	**CP-QC**	**IP-SC**	**CP-SC**	**IP-CC**	**CP-CC**
0.02	11.96 (4.03)	**78.81** (2.98)	22.57 (9.14)	15.02 (5.8)	20.91 (8.57)	15.46 (5.4)	18.96 (7.82)	13.79 (5.43)
0.05	19.35 (6.62)	**84.87** (1.63)	61.81 (7.17)	37.24 (7.55)	58.56 (7.5)	31.54 (9.11)	52.93 (7.74)	24.84 (7.85)
0.10	42.87 (7.4)	**87.13** (1.12)	81.57 (3.51)	60.04 (7.23)	80.78 (3.71)	54.66 (7.07)	78.93 (4.26)	50.4 (6.7)
0.20	68.58 (4.38)	87.61 (0.94)	**89.12** (1.5)	85.79 (2.17)	89.06 (1.5)	82.83 (2.57)	88.82 (1.47)	79.01 (3.19)
0.30	82.1 (2.02)	87.58 (0.74)	91.74 (0.87)	90.98 (1.02)	**91.75** (0.87)	89.44 (1.15)	91.73 (0.88)	87.57 (1.28)
0.50	88.3 (1.11)	87.85 (0.78)	93.87 (0.71)	93.39 (0.81)	**93.89** (0.7)	92.45 (0.86)	**93.89** (0.71)	91.37 (0.92)
0.80	89.73 (1.43)	87.88 (1.42)	**94.98** (0.99)	94.33 (1.13)	94.96 (0.98)	93.3 (1.2)	94.91 (0.96)	92.18 (1.21)

Table 3: Accuracy of various SSL methods on the digits dataset for the q-experiment with $q = 3$. We pick $\varepsilon^{(\ell)} = 100^{2-\ell}$ for $1 \leq \ell \leq 5$. Proposed methods are in bold.

RATE	LAPLACE	POISSON	WNLL	PROPERLY	p-LAP	RW	CK	**IP-VQC (2)**	**IP-VQC (3)**
0.02	12.20 (4.75)	**79.00** (2.75)	67.07 (6.07)	78.29 (3.14)	77.83 (3.23)	30.17 (11.33)	60.00 (4.17)	25.16 (9.35)	24.25 (9.65)
0.05	20.42 (7.03)	**84.61** (1.72)	69.20 (4.38)	83.11 (2.08)	82.50 (2.19)	32.00 (5.96)	66.19 (3.73)	62.69 (6.84)	61.96 (6.85)
0.10	41.62 (6.59)	86.73 (1.36)	80.73 (3.07)	**87.67** (1.45)	87.45 (1.51)	31.95 (5.56)	71.98 (2.73)	81.51 (3.66)	81.25 (3.61)
0.20	68.47 (4.79)	87.61 (0.99)	86.21 (1.53)	89.04 (0.97)	88.93 (1.00)	40.94 (4.75)	78.25 (1.53)	**89.49** (1.09)	89.41 (1.10)
0.30	82.17 (2.32)	87.62 (0.80)	88.00 (1.20)	89.81 (0.87)	89.74 (0.89)	44.89 (5.34)	82.11 (0.81)	**91.83** (0.86)	91.79 (0.83)
0.50	88.18 (1.00)	87.84 (0.96)	89.04 (1.00)	89.98 (1.00)	89.94 (0.99)	37.33 (2.51)	85.67 (0.98)	**93.79** (0.91)	93.77 (0.90)
0.80	89.65 (1.49)	87.88 (1.40)	89.68 (1.45)	89.97 (1.42)	89.97 (1.41)	33.93 (1.16)	88.34 (1.39)	94.91 (1.01)	**94.93** (1.00)

Table 4: Accuracy of various SSL methods on the digits dataset for the j-experiment with $j = 2$. We pick $\varepsilon^{(\ell)} = 100^{2-\ell}$ for $1 \leq \ell \leq 5$. Proposed methods are in bold.

RATE	LAPLACE	POISSON	**IP-QC**	**CP-QC**	**IP-SC**	**CP-SC**	**IP-CC**	**CP-CC**
1	58.08 (8.37)	57.12 (7.32)	**60.98** (7.28)	59.25 (7.54)	59.73 (7.89)	59.00 (7.85)	58.81 (8.09)	58.67 (8.07)
2	66.85 (5.49)	57.32 (6.44)	**67.75** (5.42)	67.45 (5.44)	67.26 (5.60)	67.32 (5.46)	66.85 (5.77)	67.22 (5.46)
5	73.46 (2.31)	56.83 (5.31)	73.59 (2.36)	**73.65** (2.35)	73.61 (2.42)	73.63 (2.34)	73.58 (2.48)	73.59 (2.27)
10	75.86 (1.82)	56.08 (5.31)	76.09 (1.88)	**76.21** (1.81)	76.15 (1.84)	76.14 (1.83)	76.15 (1.84)	76.06 (1.83)
20	77.61 (1.15)	56.20 (4.25)	**78.52** (1.51)	78.14 (1.18)	78.42 (1.43)	78.02 (1.17)	78.26 (1.31)	77.87 (1.15)
50	79.60 (0.88)	56.44 (3.93)	**80.95** (0.91)	80.37 (0.89)	80.83 (0.94)	80.18 (0.90)	80.64 (0.93)	80.00 (0.90)
100	80.86 (0.57)	56.06 (2.98)	**82.47** (0.70)	81.82 (0.56)	82.33 (0.62)	81.61 (0.56)	82.10 (0.61)	81.35 (0.55)

Table 5: Accuracy of various SSL methods on the Salinas A dataset for the q-experiment with $q = 3$. We pick $k^{(1)} = 50$, $k^{(2)} = 30$, $k^{(3)} = 20$ and $k^{(4)} = 10$. Proposed methods are in bold.

Rate	Laplace	Poisson	WNLL	Properly	p-Lap	RW	CK	**IP-VQC (2)**	**IP-VQC (3)**
1	59.28 (8.54)	58.31 (6.46)	**64.13** (6.05)	64.10 (6.04)	60.26 (5.44)	63.10 (5.14)	28.50 (5.98)	61.88 (7.11)	62.23 (6.78)
2	66.82 (5.35)	56.76 (7.03)	67.54 (5.04)	67.42 (5.10)	64.65 (5.13)	66.94 (4.76)	33.05 (6.65)	67.53 (5.07)	**67.68** (5.12)
5	73.74 (2.71)	55.56 (5.89)	73.42 (3.07)	73.14 (3.15)	72.26 (3.07)	73.70 (2.60)	46.37 (5.32)	**73.94** (2.84)	73.86 (2.85)
10	75.88 (1.67)	56.49 (5.18)	75.81 (1.73)	75.32 (1.81)	74.80 (1.85)	75.98 (1.73)	55.54 (4.27)	**76.23** (1.76)	76.14 (1.81)
20	77.44 (1.37)	55.99 (4.62)	78.23 (1.40)	77.56 (1.58)	77.51 (1.61)	77.99 (1.22)	66.04 (3.10)	78.34 (1.31)	**78.40** (1.37)
50	79.58 (0.94)	56.69 (4.19)	80.87 (0.90)	80.21 (0.93)	80.36 (0.88)	79.10 (0.85)	75.21 (1.76)	80.87 (0.98)	**80.98** (1.01)
100	80.96 (0.73)	55.83 (2.75)	82.10 (0.63)	81.88 (0.66)	82.12 (0.61)	79.27 (0.70)	79.82 (1.02)	82.41 (0.72)	**82.53** (0.75)

Table 6: Accuracy of various SSL methods on the Salinas A dataset for the j-experiment with $j = 2$. We pick $k^{(1)} = 50$, $k^{(2)} = 30$, $k^{(3)} = 20$ and $k^{(4)} = 10$. Proposed methods are in bold.

Rate	Laplace	Poisson	**IP-QC**	**CP-QC**	**IP-SC**	**CP-SC**	**IP-CC**	**CP-CC**
0.02	97.07 (0.07)	96.80 (0.06)	**97.36** (0.07)	97.29 (0.08)	97.26 (0.07)	97.24 (0.07)	97.19 (0.07)	97.19 (0.07)
0.05	97.37 (0.05)	96.85 (0.04)	**97.64** (0.06)	97.59 (0.06)	97.56 (0.05)	97.54 (0.06)	97.50 (0.05)	97.48 (0.06)
0.10	97.58 (0.04)	96.85 (0.04)	**97.82** (0.04)	97.77 (0.04)	97.76 (0.04)	97.74 (0.04)	97.70 (0.04)	97.69 (0.04)
0.20	97.81 (0.04)	96.87 (0.04)	**98.01** (0.04)	97.98 (0.04)	97.97 (0.04)	97.95 (0.04)	97.92 (0.04)	97.91 (0.04)
0.30	97.92 (0.04)	96.87 (0.05)	**98.10** (0.04)	98.07 (0.04)	98.07 (0.04)	98.05 (0.04)	98.02 (0.04)	98.02 (0.05)
0.50	98.08 (0.06)	96.87 (0.08)	**98.24** (0.06)	98.21 (0.06)	98.21 (0.06)	98.19 (0.06)	98.18 (0.06)	98.17 (0.06)
0.80	98.25 (0.09)	96.90 (0.12)	**98.38** (0.09)	98.36 (0.10)	98.37 (0.09)	98.34 (0.09)	98.34 (0.09)	98.32 (0.09)

Table 7: Accuracy of various SSL methods on the MNIST dataset for the q-experiment with $q = 3$. We pick $k^{(\ell)} = 30 - (\ell - 1) \cdot 10$ for $1 \leq \ell \leq 3$. Proposed methods are in bold.

Rate	Laplace	Poisson	WNLL	Properly	p-Lap	RW	CK	**IP-VQC (2)**	**IP-VQC (3)**
0.02	97.06 (0.09)	96.79 (0.07)	96.55 (0.09)	94.76 (0.17)	94.48 (0.17)	97.15 (0.10)	95.34 (0.16)	97.31 (0.09)	**97.34** (0.09)
0.05	97.37 (0.06)	96.85 (0.05)	97.20 (0.05)	94.49 (0.12)	95.49 (0.10)	97.37 (0.07)	96.46 (0.08)	97.62 (0.05)	**97.64** (0.05)
0.10	97.59 (0.04)	96.86 (0.04)	97.58 (0.05)	95.59 (0.08)	96.88 (0.06)	97.45 (0.05)	97.18 (0.06)	97.80 (0.04)	**97.82** (0.04)
0.20	97.80 (0.04)	96.87 (0.04)	97.86 (0.04)	97.08 (0.05)	97.71 (0.04)	97.50 (0.05)	97.68 (0.04)	97.99 (0.04)	**98.00** (0.04)
0.30	97.92 (0.05)	96.87 (0.05)	97.98 (0.05)	97.61 (0.06)	97.88 (0.05)	97.51 (0.05)	97.88 (0.05)	**98.10** (0.05)	**98.10** (0.05)
0.50	98.08 (0.06)	96.86 (0.06)	98.11 (0.06)	98.01 (0.06)	98.07 (0.06)	97.51 (0.06)	98.09 (0.06)	**98.24** (0.06)	**98.24** (0.05)
0.80	98.22 (0.10)	96.87 (0.14)	98.23 (0.10)	98.22 (0.11)	98.23 (0.10)	97.52 (0.13)	98.24 (0.11)	**98.37** (0.11)	**98.37** (0.11)

Table 8: Accuracy of various SSL methods on the MNIST dataset for the j-experiment with $j = 2$. We pick $k^{(\ell)} = 30 - (\ell - 1) \cdot 10$ for $1 \leq \ell \leq 3$. Proposed methods are in bold.

6 Conclusion

We have conducted a rigorous continuum analysis of hypergraph-based semi-supervised learning, revealing that classical hypergraph models—despite their combinatorial expressiveness—converge asymptotically to first-order graph-based methods. Specifically, through both pointwise and variational convergence analyses, we established that these discrete hypergraph models approach weighted Sobolev $\mathrm{W}^{1,p}$ regularization in the limit. Crucially, our results identify a previously undocumented discrepancy between the operator obtained via pointwise convergence and that derived from the variational limit. This divergence underscores the importance of considering both modes of convergence when analyzing discrete-to-continuum limits in graph- and hypergraph-based learning. Furthermore, we demonstrated that the transition between meaningful regularization and trivial smoothing is sensitive to the parameters chosen in the hypergraph construction.

To address the inherent limitations of classical hypergraph approaches, we introduced HOHL—a variational model that enforces higher-order smoothness by penalizing powers of graph Laplacians at multiple scales derived from the hypergraph structure. Our theoretical results characterize the well- and ill-posedness of HOHL, and show that it converges to genuinely higher-order Sobolev-type energies. Additionally, we argued that for point clouds embedded in a metric space, multiscale Laplacian learning is an appropriate surrogate for HOHL, thereby providing a principled theoretical foundation for this latter class of methods. Extensions to more general hypergraph settings, along with algorithmic and computational developments, are discussed in [92].

Empirical evaluations confirm the practical effectiveness of HOHL: by leveraging multiscale structure and higher-order regularization, the model achieves strong performance across a range of standard SSL benchmarks.

By analyzing hypergraph learning within a unified continuum framework, our work also offers a systematic classification of regularization-based SSL algorithms (see Figure 2) and opens new avenues for principled model design grounded in asymptotic analysis. Promising future directions include extending our framework to additional hypergraph-based methods [30, 34, 52, 62, 76].

Declarations

Data Availability Statement All data analyzed in this study are publicly available from established open-access repositories, and all data sources are explicitly referenced in the manuscript. No proprietary or restricted data were used.

Funding AW and AB were supported in part by NSF grant DMS-2152717. MT acknowledges the support of the EPSRC Mathematical and Foundations of Artificial Intelligence Probabilistic AI Hub (grant agreement EP/Y007174/1), the Leverhulme Trust through the Project Award "Robust Learning: Uncertainty Quantification, Sensitivity and Stability" (grant agreement RPG-2024-051) and the NHSBT award 177PATH25 "Harnessing Computational Genomics to Optimise Blood Transfusion Safety and Efficacy".

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Salinas hyperspectral dataset. `http://www.ehu.eus/ccwintco/index.php?Title=Hyperspectral_Remote_Sensing_Scenes`, 2008. from the Computational Intelligence Group, University of the Basque Country.

[2] Sami Abu-El-Haija, Bryan Perozzi, Amol Kapoor, Nazanin Alipourfard, Kristina Lerman, Hrayr Harutyunyan, Greg Ver Steeg, and Aram Galstyan. Mixhop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97 of *Proceedings of Machine Learning Research*, page 21-29. PMLR, 2019.

[3] Sameer Agarwal, Kristin Branson, and Serge Belongie. Higher order learning with graphs. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, page 17–24, New York, NY, USA, 2006. Association for Computing Machinery.

[4] Ethem Alpaydin and Cenk Kaynak. Optical Recognition of Handwritten Digits. UCI Machine Learning Repository, 1998. DOI: https://doi.org/10.24432/C50P49.

[5] Mikhail Belkin and Partha Niyogi. Using manifold structure for partially labelled classification. In *Advances in Neural Information Processing Systems*, pages 953–960, 2002.

[6] Mikhail Belkin and Partha Niyogi. Semi-supervised learning on Riemannian manifolds. *Machine Learning*, 56(1):209–239, 2004.

[7] Mikhail Belkin and Partha Niyogi. Convergence of Laplacian eigenmaps. In *Advances in Neural Information Processing Systems*, 2007.

[8] Andrea L. Bertozzi and Arjuna Flenner. Diffuse interface models on graphs for classification of high dimensional data. *SIAM Review*, 58(2):293–328, 2016.

[9] Maciej Besta, Florian Scheidl, Lukas Gianinazzi, Grzegorz Kwasniewski, Shachar Klaiman, Jürgen Müller, and Torsten Hoefler. Demystifying higher-order graph neural networks, 2024.

[10] Jean Bourgain, Haim Brezis, and Petru Mironescu. Another look at Sobolev spaces. In *Optimal Control and Partial Differential Equations*, pages 439–455, 2001.

[11] Andrea Braides. *Γ-convergence for Beginners*. Oxford University Press, 2002.

[12] Wlodzimierz Bryc. *The Normal Distribution: Characterizations with Applications*, volume 100 of *Lecture Notes in Statistics*. Springer New York, New York, NY, USA, 1 edition, 1995.

[13] Leon Bungert, Jeff Calder, Max Mihailescu, Kodjo Houssou, and Amber Yuan. Convergence rates for Poisson learning to a Poisson equation with measure data, 2024.

[14] Leon Bungert, Jeff Calder, and Tim Roith. Uniform convergence rates for lipschitz learning on graphs. *IMA Journal of Numerical Analysis*, 43(4):2445–2495, 09 2022.

[15] Jeff Calder. The game theoretic p-Laplacian and semi-supervised learning with few labels. *Nonlinearity*, 32(1):301, 2018.

[16] Jeff Calder. Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data. *SIAM Journal on Mathematics of Data Science*, 1(4):780–812, 2019.

[17] Jeff Calder, Brendan Cook, Matthew Thorpe, and Dejan Slepčev. Poisson learning: Graph based semi-supervised learning at very low label rates. In *Proceedings of the International Conference on Machine Learning*, pages 1283–1293, 2020.

[18] Jeff Calder and Nicolás García Trillos. Improved spectral convergence rates for graph Laplacians on ε-graphs and $k - nn$ graphs. *Applied and Computational Harmonic Analysis*, 60:123–175, 2022.

[19] Jeff Calder and Dejan Slepčev. Properly-weighted graph Laplacian for semi-supervised learning. *Applied Mathematics & Optimization*, 82(3):1111–1159, 2020.

[20] Jeff Calder, Dejan Slepčev, and Matthew Thorpe. Rates of convergence for Laplacian semi-supervised learning with low labeling rates. *Research in the Mathematical Sciences*, 10(1):10, 2023.

[21] Marco Caroccia, Antonin Chambolle, and Dejan Slepčev. Mumford–Shah functionals on graphs and their asymptotics. *Nonlinearity*, 33(8):3846–3888, jun 2020.

[22] Wu Chen, Qiuping Jiang, Wei Zhou, Long Xu, and Weisi Lin. Dynamic hypergraph convolutional network for no-reference point cloud quality assessment. *IEEE Trans. Cir. and Sys. for Video Technol.*, 34(10_Part_2):10479–10493, October 2024.

[23] Uthsav Chitra and Benjamin Raphael. Random walks on hypergraphs with edge-dependent vertex weights. In *International conference on machine learning*, pages 1172–1181. PMLR, 2019.

[24] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

[25] Riccardo Cristoferi and Matthew Thorpe. Large data limit for a phase transition model with the p-Laplacian on point clouds. *European Journal of Applied Mathematics*, 31(2):185–231, 2020.

[26] Matthew Dunlop, Dejan Slepcev, Andrew Stuart, and Matthew Thorpe. Large data and zero noise limits of graph-based semi-supervised learning algorithms. *Applied and Computational Harmonic Analysis*, 49(2):655–697, 2020.

[27] Ahmed El Alaoui, Xiang Cheng, Aaditya Ramdas, Martin J. Wainwright, and Michael I. Jordan. Asymptotic behavior of ℓ_p-based Laplacian regularization in semi-supervised learning. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 879–906, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR.

[28] Imad El Bouchairi, Jalal Fadili, and Abderrahim Elmoataz. Continuum limit of p-Laplacian evolution problems on graphs: l^q graphons and sparse graphs. *ESAIM: Mathematical Modelling and Numerical Analysis*, 2023. arXiv 2010.08697.

[29] K.-T. Fang, S. Kotz, and W. Ng, K.˙ *Symmetric Multivariate and Related Distributions*. Chapman & Hall/CRC, New York, 1 edition, 1990.

[30] Ariane Fazeny, Daniel Tenbrinck, and Martin Burger. Hypergraph p-Laplacians, scale spaces, and information flow in networks. In Luca Calatroni, Marco Donatelli, Serena Morigi, Marco Prato, and Matteo Santacesaria, editors, *Scale Space and Variational Methods in Computer Vision*, pages 677–690, Cham, 2023. Springer International Publishing.

[31] Ronald Fisher. Iris. UCI Machine Learning Repository, 1988. DOI: https://doi.org/10.24432/C56C76.

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

[33] Yue Gao, Meng Wang, Zheng-Jun Zha, Jialie Shen, Xuelong Li, and Xindong Wu. Visual-textual joint relevance learning for tag-based social image search. *IEEE Transactions on Image Processing*, 22(1):363–376, 2013.

[34] Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2548–2566, 2022.

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

[36] Nicolás García Trillos and Ryan Murray. A new analytical approach to consistency and overfitting in regularized empirical risk minimization. *European Journal of Applied Mathematics*, 28(6):886–921, 2017.

[37] Nicolás García Trillos, Ryan Murray, and Matthew Thorpe. From graph cuts to isoperimetric inequalities: Convergence rates of Cheeger cuts on data clouds. *Archive for Rational Mechanics and Analysis*, 244(3):541–598, 2022.

[38] Nicolás García Trillos, Ryan Murray, and Matthew Thorpe. Rates of convergence for regression with the graph poly-Laplacian. *Sampling Theory, Signal Processing, and Data Analysis*, 21(2):35, 2023.

[39] Nicolás García Trillos and Dejan Slepčev. Continuum limit of total variation on point clouds. *Archive for Rational Mechanics and Analysis*, 220(1):193–241, 2016.

[40] Nicolás García Trillos and Dejan Slepčev. A variational approach to the consistency of spectral clustering. *Applied and Computational Harmonic Analysis*, 45(2):239–281, 2018.

[41] Nicolás García Trillos, Dejan Slepčev, and James Von Brecht. Estimating perimeter using graph cuts. *Advances in Applied Probability*, 49(4):1067–1090, 2017.

[42] Nicolás García Trillos, Dejan Slepčev, James von Brecht, Thomas Laurent, and Xavier Bresson. Consistency of Cheeger and ratio graph cuts. *Journal of Machine Learning Research*, 17(181):1–46, 2016.

[43] Nicolás García Trillos and Dejan Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, 2015.

[44] Johannes Gasteiger, Stefan Weiß enberger, and Stephan Günnemann. Diffusion improves graph learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[45] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral partitioning of uniform hypergraphs under planted partition model. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[46] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral hypergraph partitioning under planted partition model. *The Annals of Statistics*, 45(1):289–315, 2017.

[47] Evarist Giné and Vladimir Koltschinskii. *Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results*, volume 51 of *IMS Lecture Notes Monographs Series*, pages 238–259. Institute of Mathematical Statistics, 2006.

[48] Ashish Goel, Sanatan Rai, and Bhaskar Krishnamachari. Monotone properties of random geometric graphs have sharp thresholds. *The Annals of Applied Probability*, 15:2535–2552, 2005.

[49] Matthias Hein. Uniform convergence of adaptive graph-based regularization. In *Proceedings of the Conference on Learning Theory*, pages 50–64, 2006.

[50] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. Graph Laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8(6), 2007.

[51] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. From graphs to manifolds – weak and strong pointwise consistency of graph Laplacians. In *Proceedings of the Conference on Learning Theory*, pages 470–485, 2005.

[52] Matthias Hein, Simon Setzer, Leonardo Jost, and Syama Sundar Rangapuram. The total variation on hypergraphs - learning on hypergraphs revisited. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 2427–2435, Red Hook, NY, USA, 2013. Curran Associates Inc.

[53] Jürgen Jost and Raffaella Mulas. Hypergraph Laplace operators for chemical reaction networks. *Advances in Mathematics*, 351:870–896, 2019.

[54] Jürgen Jost, Raffaella Mulas, and Dong Zhang. p-Laplace operators for oriented hypergraphs. *Vietnam Journal of Mathematics*, 50(2):323–358, 2022.

[55] Wei Ju, Siyu Yi, Yifan Wang, Qingqing Long, Junyu Luo, Zhiping Xiao, and Ming Zhang. A survey of data-efficient graph learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, IJCAI '24, 2024.

[56] Alexander Jung, Alfred O Hero III, Alexandru Mara, and Saeed Jahromi. Semi-supervised learning via sparse label propagation. *arXiv preprint arXiv:1612.01414*, 2016.

[57] Samuel Kotz, N. Balakrishnan, and Norman L. Johnson. *Continuous Multivariate Distributions, Volume 1: Models and Applications*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ, USA, 2 edition, 2000.

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

[59] 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.

[60] John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer Science+Business Media, New York, NY, 2012.

[61] Giovanni Leoni. *A First Course in Sobolev Spaces*. Graduate studies in mathematics. American Mathematical Society, 2017.

[62] Pan Li and Olgica Milenkovic. Submodular hypergraphs: p-Laplacians, Cheeger inequalities and spectral clustering. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3014–3023. PMLR, 10–15 Jul 2018.

[63] Xiaoyi Mai. A random matrix analysis and improvement of semi-supervised learning for large dimensional data. *Journal of Machine Learning Research*, 19(79):1–27, 2018.

[64] Colin McDiarmid. On the method of bounded differences. In J. Siemons, editor, *Surveys in Combinatorics, 1989*, volume 141 of *London Mathematical Society Lecture Note Series*, pages 148–188. Cambridge University Press, Cambridge, 1989.

[65] Ekaterina Merkurjev, Duc Duy Nguyen, and Guo-Wei Wei. Multiscale Laplacian learning. *Applied Intelligence*, 53(12):15727–15746, nov 2022.

[66] Federico Monti, Karl Otness, and Michael M. Bronstein. Motifnet: A motif-based graph convolutional network for directed graphs. In *2018 IEEE Data Science Workshop (DSW)*, pages 225–228, 2018.

[67] Raffaella Mulas, Christian Kuehn, Tobias Böhle, and Jürgen Jost. Random walks and Laplacians on hypergraphs: When do they match? *Discrete Applied Mathematics*, 317:26–41, 2022.

[68] Boaz Nadler, Nathan Srebro, and Xueyuan Zhou. Semi-supervised learning with the graph Laplacian: The limit of infinite unlabelled data. In *Advances in Neural Information Processing Systems*, pages 1330–1338, 2009.

[69] Leonie Neuhäuser, Renaud Lambiotte, and Michael T. Schaub. Consensus dynamics and opinion formation on hypergraphs. In Federico Battiston and Giovanni Petri, editors, *Higher-Order Systems*, pages 347–376. Springer International Publishing, Cham, 2022.

[70] Braxton Osting and Todd Harry Reeb. Consistency of Dirichlet partitions. *SIAM Journal on Mathematical Analysis*, 49(5):4251–4274, 2017.

[71] Bruno Pelletier and Pierre Pudlo. Operator norm convergence of spectral clustering on level sets. *Journal of Machine Learning Research*, 12(12):385–416, 2011.

[72] Mathew D. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[73] Mihai Pirvu, Alina Marcu, Maria Alexandra Dobrescu, Ahmed Nabil Belbachir, and Marius Leordeanu. Multi-task hypergraphs for semi-supervised learning using earth observations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 3404–3414, October 2023.

[74] Augusto 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.

[75] Tim Roith and Leon Bungert. Continuum limit of Lipschitz learning on graphs. *Foundations of Computational Mathematics*, pages 1–39, 2022.

[76] Shota Saito, Danilo P Mandic, and Hideyuki Suzuki. Hypergraph p-Laplacian: a differential geometry view. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18. AAAI Press, 2018.

[77] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser Basel, 2015.

[78] Kehan Shi and Martin Burger. Hypergraph p-Laplacian equations for data interpolation and semi-supervised learning. *Journal of Scientific Computing*, 103(3):93, 2025.

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

[80] Zuoqiang Shi, Stanley Osher, and Wei Zhu. Generalization of the weighted nonlocal Laplacian in low dimensional manifold model. *Journal of Scientific Computing*, 75(2):638–656, 2018.

[81] Zuoqiang Shi, Bao Wang, and Stanley J Osher. Error estimation of weighted nonlocal Laplacian on random point cloud. *arXiv preprint arXiv:1809.08622*, 2018.

[82] Amit Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21:128–134, 2006.

[83] Amit Singer and Hau-Tieng Wu. Spectral convergence of the connection Laplacian from random samples. *Information and Inference: A Journal of the IMA*, 6(1):58–123, 12 2016.

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

[85] Matthew Thorpe and Florian Theil. Asymptotic analysis of the Ginzburg–Landau functional on point clouds. *Proceedings of the Royal Society of Edinburgh: Section A Mathematics*, 149(2):387–427, 2019.

[86] Daniel Ting, Ling Huang, and Michael I. Jordan. An analysis of the convergence of graph Laplacians. In *Proceedings of the International Conference on Machine Learning*, pages 1079–1086, 2010.

[87] Yves van Gennip and Andrea Bertozzi. Gamma-convergence of graph Ginzburg–Landau functionals. *Advances in Differential Equations*, 17(11–12):1115–1180, 2012.

[88] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer-Verlag Berlin Heidelberg, 2009.

[89] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 2007.

[90] Ulrike von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, 2008.

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

[92] Adrien Weihs, Andrea Bertozzi, and Matthew Thorpe. Higher-order regularization learning on hypergraphs, 2025.

[93] Adrien Weihs, Jalal Fadili, and Matthew Thorpe. Discrete-to-continuum rates of convergence for nonlocal p-Laplacian evolution problems. *Information and Inference: A Journal of the IMA*, 13(4):iaae031, 11 2024.

[94] Adrien Weihs and Matthew Thorpe. Consistency of fractional graph-Laplacian regularization in semisupervised learning with finite labels. *SIAM Journal on Mathematical Analysis*, 56(4):4253–4295, 2024.

[95] Feng Xia, Ke Sun, Shuo Yu, Abdul Aziz, Liangtian Wan, Shirui Pan, and Huan Liu. Graph learning: A survey. *IEEE Transactions on Artificial Intelligence*, 2(2):109–127, 2021.

[96] Chaoqi Yang, Ruijie Wang, Shuochao Yao, and Tarek Abdelzaher. Semi-supervised hypergraph node classification on hypergraph line expansion. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, page 2352–2361, New York, NY, USA, 2022. Association for Computing Machinery.

[97] Damián H. Zanette. Beyond networks: opinion formation in triplet-based populations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1901):3311–3319, 2009.

[98] Zizhao Zhang, Haojie Lin, and Yue Gao. Dynamic hypergraph structure learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, page 3162–3169. AAAI Press, 2018.

[99] Dengyong Zhou, Thomas Hofmann, and Bernhard Schölkopf. Semi-supervised learning on directed graphs. In L. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2004.

[100] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press, 2006.

[101] Dengyong Zhou and Bernhard Schölkopf. Learning from labeled and unlabeled data using random walks. In Carl Edward Rasmussen, Heinrich H. Bülthoff, Bernhard Schölkopf, and Martin A. Giese, editors, *Pattern Recognition*, pages 237–244, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[102] Xueyuan Zhou and Mikhail Belkin. Semi-supervised learning by higher order regularization. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 892–900, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.

[103] Xianjin Zhu, Zoubin Ghahramani, and John Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *Proceedings of the International Conference on Machine Learning*, 2003.

7 Complete numerical experiments

In this section, we present the complete numerical experiments of Section 5. We refer to Table 2 for a review of the terminology used throughout the experiments.

Table 9: Accuracy of various SSL methods on the iris dataset. We pick $\varepsilon^{(k)} = 2^{3-k}$ for $1 \leq k \leq 5$. Proposed methods are in bold.

q	rate	Laplace	Poisson	FL ($s=2$)	FL ($s=3$)	**IP-QC**	**CP-QC**	**IP-SC**	**CP-SC**	**IP-CC**	**CP-CC**
2	0.02	64.56 (15.0)	**79.39** (8.83)	70.73 (7.46)	71.54 (7.76)	75.56 (10.01)	70.54 (10.84)	75.02 (10.0)	68.75 (12.77)	73.89 (10.82)	68.29 (12.5)
	0.05	73.15 (9.26)	80.54 (6.05)	74.47 (9.26)	76.19 (9.13)	**82.26** (8.98)	74.65 (9.73)	81.62 (9.19)	73.94 (9.66)	80.46 (9.69)	73.67 (9.78)
	0.10	81.05 (8.94)	80.58 (3.25)	82.55 (8.21)	84.37 (7.27)	**90.39** (3.84)	85.41 (7.95)	90.29 (4.23)	84.44 (8.47)	89.93 (4.75)	83.36 (8.81)
	0.20	87.43 (5.88)	80.33 (2.3)	87.91 (5.25)	88.74 (4.14)	92.48 (2.49)	90.53 (3.9)	92.52 (2.59)	89.96 (4.38)	**92.54** (2.69)	89.22 (4.92)
	0.30	90.82 (3.2)	79.66 (2.18)	90.88 (2.87)	91.05 (2.53)	93.57 (2.48)	92.51 (2.82)	93.57 (2.46)	92.36 (3.01)	**93.61** (2.52)	92.01 (3.02)
	0.50	91.81 (2.71)	79.59 (2.34)	91.67 (2.9)	91.8 (2.77)	**94.92** (2.3)	93.52 (2.77)	94.89 (2.31)	93.33 (2.73)	94.85 (2.32)	93.04 (2.75)
	0.80	92.2 (4.67)	79.33 (5.44)	92.2 (4.35)	92.57 (4.44)	95.5 (3.49)	94.47 (3.52)	**95.53** (3.49)	93.93 (3.49)	**95.53** (3.49)	93.67 (3.5)
3	0.02	64.56 (15.0)	**79.39** (8.83)	70.73 (7.46)	71.54 (7.76)	76.52 (10.76)	71.27 (11.86)	75.99 (10.48)	69.29 (13.27)	74.5 (11.18)	68.47 (12.74)
	0.05	73.15 (9.26)	80.54 (6.05)	74.47 (9.26)	76.19 (9.13)	**84.78** (9.8)	75.72 (10.08)	83.78 (10.0)	74.33 (9.99)	82.26 (10.21)	73.53 (9.82)
	0.10	81.05 (8.94)	80.58 (3.25)	82.55 (8.21)	84.37 (7.27)	**91.99** (3.32)	87.44 (7.5)	91.48 (3.54)	85.9 (8.5)	91.04 (4.06)	84.23 (8.92)
	0.20	87.43 (5.88)	80.33 (2.3)	87.91 (5.25)	88.74 (4.14)	**94.04** (2.49)	92.03 (3.19)	93.79 (2.51)	91.12 (3.9)	93.53 (2.61)	90.09 (4.69)
	0.30	90.82 (3.2)	79.66 (2.18)	90.88 (2.87)	91.05 (2.53)	**95.52** (1.86)	94.04 (2.64)	95.3 (1.95)	93.32 (2.87)	94.95 (2.22)	92.59 (3.04)
	0.50	91.81 (2.71)	79.59 (2.34)	91.67 (2.9)	91.8 (2.77)	**95.89** (1.86)	94.97 (2.45)	95.87 (1.9)	94.41 (2.62)	**95.89** (1.9)	93.77 (2.8)
	0.80	92.2 (4.67)	79.33 (5.44)	92.2 (4.35)	92.57 (4.44)	96.17 (3.23)	95.73 (3.67)	**96.27** (3.29)	95.43 (3.72)	96.13 (3.37)	94.57 (3.72)
4	0.02	64.56 (15.0)	**79.39** (8.83)	70.73 (7.46)	71.54 (7.76)	76.53 (10.81)	71.46 (12.53)	75.97 (10.56)	69.34 (13.33)	74.46 (11.17)	68.41 (12.69)
	0.05	73.15 (9.26)	80.54 (6.05)	74.47 (9.26)	76.19 (9.13)	**84.88** (9.8)	76.15 (10.27)	83.81 (9.98)	74.42 (9.92)	82.33 (10.26)	73.52 (9.87)
	0.10	81.05 (8.94)	80.58 (3.25)	82.55 (8.21)	84.37 (7.27)	**92.01** (3.3)	87.63 (7.6)	91.47 (3.54)	86.04 (8.52)	91.07 (4.03)	84.27 (8.96)
	0.20	87.43 (5.88)	80.33 (2.3)	87.91 (5.25)	88.74 (4.14)	**94.08** (2.47)	92.4 (3.25)	93.8 (2.52)	91.32 (3.98)	93.54 (2.62)	90.22 (4.6)
	0.30	90.82 (3.2)	79.66 (2.18)	90.88 (2.87)	91.05 (2.53)	**95.53** (1.84)	94.46 (2.68)	95.33 (1.97)	93.58 (2.87)	94.95 (2.22)	92.73 (3.06)
	0.50	91.81 (2.71)	79.59 (2.34)	91.67 (2.9)	91.8 (2.77)	95.88 (1.84)	95.51 (2.33)	95.87 (1.87)	94.79 (2.64)	**95.92** (1.89)	93.95 (2.71)
	0.80	92.2 (4.67)	79.33 (5.44)	92.2 (4.35)	92.57 (4.44)	96.17 (3.23)	95.9 (3.6)	**96.23** (3.27)	95.83 (3.65)	96.17 (3.36)	94.9 (3.8)
5	0.02	64.56 (15.0)	**79.39** (8.83)	70.73 (7.46)	71.54 (7.76)	76.6 (10.82)	71.38 (12.49)	75.98 (10.56)	69.14 (13.28)	74.46 (11.16)	68.37 (12.68)
	0.05	73.15 (9.26)	80.54 (6.05)	74.47 (9.26)	76.19 (9.13)	**84.89** (9.81)	76.31 (10.26)	83.83 (9.98)	74.44 (9.86)	82.33 (10.26)	73.53 (9.83)
	0.10	81.05 (8.94)	80.58 (3.25)	82.55 (8.21)	84.37 (7.27)	**92.02** (3.3)	87.59 (7.53)	91.47 (3.54)	85.98 (8.48)	91.07 (4.03)	84.29 (8.96)
	0.20	87.43 (5.88)	80.33 (2.3)	87.91 (5.25)	88.74 (4.14)	**94.08** (2.44)	92.49 (3.27)	93.81 (2.52)	91.34 (4.0)	93.54 (2.62)	90.27 (4.63)
	0.30	90.82 (3.2)	79.66 (2.18)	90.88 (2.87)	91.05 (2.53)	**95.53** (1.84)	94.53 (2.69)	95.33 (1.97)	93.65 (2.87)	94.95 (2.22)	92.76 (3.07)
	0.50	91.81 (2.71)	79.59 (2.34)	91.67 (2.9)	91.8 (2.77)	95.88 (1.84)	95.63 (2.28)	95.87 (1.87)	94.85 (2.61)	**95.92** (1.89)	93.93 (2.71)
	0.80	92.2 (4.67)	79.33 (5.44)	92.2 (4.35)	92.57 (4.44)	96.17 (3.23)	96.13 (3.47)	**96.23** (3.27)	95.87 (3.61)	96.17 (3.36)	94.9 (3.8)

Table 10: Accuracy of various SSL methods on the iris dataset. We pick $\varepsilon^{(\ell)} = 2^{3-\ell}$ for $1 \leq \ell \leq 5$. Proposed methods are in bold.

j	rate	Laplace	Poisson	FL ($s=2$)	FL ($s=3$)	WNLL	p-Lap	RW	CK	SLP	**IP-VQC (2)**	**IP-VQC (3)**
1	0.02	61.86 (13.18)	82.17 (8.39)	70.41 (6.98)	71.54 (7.29)	82.59 (8.64)	**86.56**(8.18)	77.13 (9.09)	76.24 (8.18)	32.05 (6.24)	59.76 (13.85)	53.71 (15.23)
	0.05	71.06 (7.31)	81.55 (5.92)	72.73 (8.16)	74.31 (7.75)	84.42 (7.85)	**88.54** (6.52)	78.38 (9.14)	79.48 (7.88)	34.97 (7.18)	70.28 (9.06)	69.32 (8.69)
	0.10	81.07 (8.72)	80.23 (3.19)	82.76 (7.99)	84.78 (7.04)	89.34 (4.42)	**91.61** (2.74)	82.34 (7.84)	85.9 (4.5)	46.67 (19.38)	81.03 (10.58)	80.19 (10.61)
	0.20	87.58 (4.63)	80.12 (2.62)	88.2 (4.11)	89.09 (3.37)	90.57 (2.57)	**91.68** (2.13)	84.25 (6.29)	89.39 (2.82)	64.05 (12.29)	90.46 (3.6)	90.31 (3.74)
	0.30	90.12 (3.0)	79.69 (2.11)	90.21 (2.97)	90.36 (2.42)	91.11 (2.25)	92.06 (2.28)	86.19 (5.72)	91.26 (2.54)	58.99 (15.82)	92.1 (3.21)	**92.29** (3.26)
	0.50	91.37 (3.23)	79.44 (2.37)	91.36 (3.01)	91.41 (3.03)	91.4 (3.23)	92.07 (3.09)	88.52 (4.53)	92.29 (2.53)	56.39 (16.53)	93.4 (3.22)	**93.88** (3.11)
	0.80	92.0 (4.42)	78.17 (4.4)	91.67 (4.3)	91.97 (4.16)	92.07 (4.36)	92.5 (4.14)	88.97 (4.89)	92.97 (4.37)	89.7 (4.88)	94.5 (3.86)	**95.37** (3.51)
2	0.02	61.86 (13.18)	82.17 (8.39)	70.41 (6.98)	71.54 (7.29)	82.59 (8.64)	**86.56** (8.18)	77.13 (9.09)	76.24 (8.18)	32.05 (6.24)	67.74 (12.14)	67.59 (12.34)
	0.05	71.06 (7.31)	81.55 (5.92)	72.73 (8.16)	74.31 (7.75)	84.42 (7.85)	**88.54** (6.52)	78.38 (9.14)	79.48 (7.88)	34.97 (7.18)	72.77 (9.79)	72.53 (9.82)
	0.10	81.07 (8.72)	80.23 (3.19)	82.76 (7.99)	84.78 (7.04)	89.34 (4.42)	**91.61** (2.74)	82.34 (7.84)	85.9 (4.5)	46.67 (19.38)	83.77 (10.48)	83.66 (10.64)
	0.20	87.58 (4.63)	80.12 (2.62)	88.2 (4.11)	89.09 (3.37)	90.57 (2.57)	**91.68** (2.13)	84.25 (6.29)	89.39 (2.82)	64.05 (12.29)	91.52 (3.07)	91.57 (3.1)
	0.30	90.12 (3.0)	79.69 (2.11)	90.21 (2.97)	90.36 (2.42)	91.11 (2.25)	92.06 (2.28)	86.19 (5.72)	91.26 (2.54)	58.99 (15.82)	92.82 (3.06)	**92.93** (3.06)
	0.50	91.37 (3.23)	79.44 (2.37)	91.36 (3.01)	91.41 (3.03)	91.4 (3.23)	92.07 (3.09)	88.52 (4.53)	92.29 (2.53)	56.39 (16.53)	93.91 (2.99)	**94.24** (2.92)
	0.80	92.0 (4.42)	78.17 (4.4)	91.67 (4.3)	91.97 (4.16)	92.07 (4.36)	92.5 (4.14)	88.97 (4.89)	92.97 (4.37)	89.7 (4.88)	95.47 (3.75)	**95.8** (3.44)
3	0.02	61.86 (13.18)	82.17 (8.39)	70.41 (6.98)	71.54 (7.29)	82.59 (8.64)	**86.56** (8.18)	77.13 (9.09)	76.24 (8.18)	32.05 (6.24)	70.62 (9.65)	70.09 (10.95)
	0.05	71.06 (7.31)	81.55 (5.92)	72.73 (8.16)	74.31 (7.75)	84.42 (7.85)	**88.54** (6.52)	78.38 (9.14)	79.48 (7.88)	34.97 (7.18)	73.8 (10.51)	73.49 (10.33)
	0.10	81.07 (8.72)	80.23 (3.19)	82.76 (7.99)	84.78 (7.04)	89.34 (4.42)	**91.61** (2.74)	82.34 (7.84)	85.9 (4.5)	46.67 (19.38)	85.27 (9.58)	85.34 (9.78)
	0.20	87.58 (4.63)	80.12 (2.62)	88.2 (4.11)	89.09 (3.37)	90.57 (2.57)	91.68 (2.13)	84.25 (6.29)	89.39 (2.82)	64.05 (12.29)	91.9 (2.9)	**92.04** (2.96)
	0.30	90.12 (3.0)	79.69 (2.11)	90.21 (2.97)	90.36 (2.42)	91.11 (2.25)	92.06 (2.28)	86.19 (5.72)	91.26 (2.54)	58.99 (15.82)	92.94 (3.0)	**93.15** (3.05)
	0.50	91.37 (3.23)	79.44 (2.37)	91.36 (3.01)	91.41 (3.03)	91.4 (3.23)	92.07 (3.09)	88.52 (4.53)	92.29 (2.53)	56.39 (16.53)	94.03 (2.99)	**94.31** (2.85)
	0.80	92.0 (4.42)	78.17 (4.4)	91.67 (4.3)	91.97 (4.16)	92.07 (4.36)	92.5 (4.14)	88.97 (4.89)	92.97 (4.37)	89.7 (4.88)	95.5 (3.71)	**95.73** (3.45)
4	0.02	61.86 (13.18)	82.17 (8.39)	70.41 (6.98)	71.54 (7.29)	82.59 (8.64)	**86.56** (8.18)	77.13 (9.09)	76.24 (8.18)	32.05 (6.24)	70.98 (9.86)	70.96 (9.8)
	0.05	71.06 (7.31)	81.55 (5.92)	72.73 (8.16)	74.31 (7.75)	84.42 (7.85)	**88.54** (6.52)	78.38 (9.14)	79.48 (7.88)	34.97 (7.18)	74.24 (10.64)	74.03 (10.44)
	0.10	81.07 (8.72)	80.23 (3.19)	82.76 (7.99)	84.78 (7.04)	89.34 (4.42)	**91.61** (2.74)	82.34 (7.84)	85.9 (4.5)	46.67 (19.38)	86.04 (9.08)	86.18 (9.05)
	0.20	87.58 (4.63)	80.12 (2.62)	88.2 (4.11)	89.09 (3.37)	90.57 (2.57)	91.68 (2.13)	84.25 (6.29)	89.39 (2.82)	64.05 (12.29)	92.09 (2.92)	**92.18** (2.91)
	0.30	90.12 (3.0)	79.69 (2.11)	90.21 (2.97)	90.36 (2.42)	91.11 (2.25)	92.06 (2.28)	86.19 (5.72)	91.26 (2.54)	58.99 (15.82)	93.04 (3.01)	**93.19** (3.04)
	0.50	91.37 (3.23)	79.44 (2.37)	91.36 (3.01)	91.41 (3.03)	91.4 (3.23)	92.07 (3.09)	88.52 (4.53)	92.29 (2.53)	56.39 (16.53)	94.08 (2.97)	**94.31** (2.9)
	0.80	92.0 (4.42)	78.17 (4.4)	91.67 (4.3)	91.97 (4.16)	92.07 (4.36)	92.5 (4.14)	88.97 (4.89)	92.97 (4.37)	89.7 (4.88)	95.53 (3.74)	**95.8** (3.47)

Table 11: Accuracy of various SSL methods on the digits dataset. We pick $\varepsilon^{(k)} = 100^{2-k}$ for $1 \leq k \leq 5$. Proposed methods are in bold.

q	rate	Laplace	Poisson	**IP-QC**	**CP-QC**	**IP-SC**	**CP-SC**	**IP-CC**	**CP-CC**
2	0.02	11.96 (4.03)	**78.81** (2.98)	24.19 (8.92)	15.81 (5.5)	21.91 (8.44)	14.88 (5.48)	19.55 (7.77)	13.44 (5.08)
	0.05	19.35 (6.62)	**84.87** (1.63)	62.35 (7.28)	34.88 (8.75)	59.01 (7.52)	29.09 (9.18)	53.38 (7.77)	23.86 (7.5)
	0.10	42.87 (7.4)	**87.13** (1.12)	81.84 (3.6)	58.25 (7.34)	80.96 (3.81)	53.24 (6.98)	79.07 (4.3)	49.71 (6.62)
	0.20	68.58 (4.38)	87.61 (0.94)	**89.21** (1.5)	84.77 (2.24)	89.11 (1.48)	81.91 (2.7)	88.86 (1.44)	78.27 (3.3)
	0.30	82.1 (2.02)	87.58 (0.74)	**91.78** (0.86)	90.13 (1.08)	**91.78** (0.88)	88.85 (1.2)	91.74 (0.89)	87.13 (1.3)
	0.50	88.3 (1.11)	87.85 (0.78)	93.87 (0.72)	92.78 (0.87)	93.87 (0.72)	92.01 (0.87)	**93.91** (0.7)	91.08 (0.93)
	0.80	89.73 (1.43)	87.88 (1.42)	**94.96** (0.98)	93.64 (1.16)	94.94 (0.97)	92.86 (1.22)	94.9 (0.96)	91.89 (1.22)
3	0.02	11.96 (4.03)	**78.81** (2.98)	22.57 (9.14)	15.02 (5.8)	20.91 (8.57)	15.46 (5.4)	18.96 (7.82)	13.79 (5.43)
	0.05	19.35 (6.62)	**84.87** (1.63)	61.81 (7.17)	37.24 (7.55)	58.56 (7.5)	31.54 (9.11)	52.93 (7.74)	24.84 (7.85)
	0.10	42.87 (7.4)	**87.13** (1.12)	81.57 (3.51)	60.04 (7.23)	80.78 (3.71)	54.66 (7.07)	78.93 (4.26)	50.4 (6.7)
	0.20	68.58 (4.38)	87.61 (0.94)	**89.12** (1.5)	85.79 (2.17)	89.06 (1.5)	82.83 (2.57)	88.82 (1.47)	79.01 (3.19)
	0.30	82.1 (2.02)	87.58 (0.74)	91.74 (0.87)	90.98 (1.02)	**91.75** (0.87)	89.44 (1.15)	91.73 (0.88)	87.57 (1.28)
	0.50	88.3 (1.11)	87.85 (0.78)	93.87 (0.71)	93.39 (0.81)	**93.89** (0.7)	92.45 (0.86)	**93.89** (0.71)	91.37 (0.92)
	0.80	89.73 (1.43)	87.88 (1.42)	**94.98** (0.99)	94.33 (1.13)	94.96 (0.98)	93.3 (1.2)	94.91 (0.96)	92.18 (1.21)
4	0.02	11.96 (4.03)	**78.81** (2.98)	22.57 (9.14)	15.03 (5.82)	20.91 (8.57)	15.46 (5.41)	18.96 (7.82)	13.79 (5.43)
	0.05	19.35 (6.62)	**84.87** (1.63)	61.81 (7.17)	37.29 (7.55)	58.56 (7.5)	31.56 (9.11)	52.93 (7.74)	24.85 (7.85)
	0.10	42.87 (7.4)	**87.13** (1.12)	81.57 (3.51)	60.09 (7.24)	80.78 (3.71)	54.68 (7.06)	78.93 (4.26)	50.41 (6.71)
	0.20	68.58 (4.38)	87.61 (0.94)	**89.12** (1.5)	85.83 (2.18)	89.06 (1.5)	82.84 (2.57)	88.82 (1.47)	79.01 (3.19)
	0.30	82.1 (2.02)	87.58 (0.74)	91.74 (0.87)	91.0 (1.02)	**91.75** (0.87)	89.45 (1.15)	91.73 (0.88)	87.57 (1.29)
	0.50	88.3 (1.11)	87.85 (0.78)	93.87 (0.71)	93.4 (0.81)	**93.89** (0.7)	92.45 (0.86)	**93.89** (0.71)	91.38 (0.92)
	0.80	89.73 (1.43)	87.88 (1.42)	**94.98** (0.99)	94.34 (1.13)	94.96 (0.98)	93.3 (1.2)	94.91 (0.96)	92.18 (1.21)
5	0.02	11.96 (4.03)	**78.81** (2.98)	22.57 (9.14)	15.03 (5.82)	20.91 (8.57)	15.46 (5.41)	18.96 (7.82)	13.79 (5.43)
	0.05	19.35 (6.62)	**84.87** (1.63)	61.81 (7.17)	37.29 (7.55)	58.56 (7.5)	31.56 (9.11)	52.93 (7.74)	24.85 (7.85)
	0.10	42.87 (7.4)	**87.13** (1.12)	81.57 (3.51)	60.09 (7.24)	80.78 (3.71)	54.68 (7.06)	78.93 (4.26)	50.41 (6.71)
	0.20	68.58 (4.38)	87.61 (0.94)	**89.12** (1.5)	85.83 (2.18)	89.06 (1.5)	82.84 (2.57)	88.82 (1.47)	79.01 (3.19)
	0.30	82.1 (2.02)	87.58 (0.74)	91.74 (0.87)	91.0 (1.02)	**91.75** (0.87)	89.45 (1.15)	91.73 (0.88)	87.57 (1.28)
	0.50	88.3 (1.11)	87.85 (0.78)	93.87 (0.71)	93.4 (0.81)	**93.89** (0.7)	92.45 (0.86)	**93.89** (0.71)	91.38 (0.92)
	0.80	89.73 (1.43)	87.88 (1.42)	**94.98** (0.99)	94.34 (1.13)	94.96 (0.98)	93.3 (1.2)	94.91 (0.96)	92.18 (1.21)

Table 12: Accuracy of various SSL methods on the digits dataset. We pick $\varepsilon^{(\ell)} = 100^{2-\ell}$ for $1 \leq \ell \leq 5$. Proposed methods are in bold.

j	rate	Laplace	Poisson	WNLL	Properly	p-Lap	RW	CK	**IP-VQC (2)**	**IP-VQC (3)**
1	0.02	12.2 (4.75)	**79.0** (2.75)	67.07 (6.07)	78.29 (3.14)	77.83 (3.23)	30.17 (11.33)	60.0 (4.17)	20.58 (8.29)	19.66 (8.71)
	0.05	20.42 (7.03)	**84.61** (1.72)	69.2 (4.38)	83.11 (2.08)	82.5 (2.19)	32.0 (5.96)	66.19 (3.73)	53.07 (7.79)	50.55 (8.44)
	0.10	41.62 (6.59)	86.73 (1.36)	80.73 (3.07)	**87.67** (1.45)	87.45 (1.51)	31.95 (5.56)	71.98 (2.73)	78.63 (4.42)	77.94 (4.46)
	0.20	68.47 (4.79)	87.61 (0.99)	86.21 (1.53)	89.04 (0.97)	88.93 (1.0)	40.94 (4.75)	78.25 (1.53)	**89.19** (1.11)	88.97 (1.1)
	0.30	82.17 (2.32)	87.62 (0.8)	88.0 (1.2)	89.81 (0.87)	89.74 (0.89)	44.89 (5.34)	82.11 (0.81)	**91.75** (0.84)	91.67 (0.84)
	0.50	88.18 (1.0)	87.84 (0.96)	89.04 (1.0)	89.98 (1.0)	89.94 (0.99)	37.33 (2.51)	85.67 (0.98)	**93.8** (0.87)	93.77 (0.86)
	0.80	89.65 (1.49)	87.88 (1.4)	89.68 (1.45)	89.97 (1.42)	89.97 (1.41)	33.93 (1.16)	88.34 (1.39)	94.86 (1.0)	**94.89** (1.02)
2	0.02	12.2 (4.75)	**79.0** (2.75)	67.07 (6.07)	78.29 (3.14)	77.83 (3.23)	30.17 (11.33)	60.0 (4.17)	25.16 (9.35)	24.25 (9.65)
	0.05	20.42 (7.03)	**84.61** (1.72)	69.2 (4.38)	83.11 (2.08)	82.5 (2.19)	32.0 (5.96)	66.19 (3.73)	62.69 (6.84)	61.96 (6.85)
	0.10	41.62 (6.59)	86.73 (1.36)	80.73 (3.07)	**87.67** (1.45)	87.45 (1.51)	31.95 (5.56)	71.98 (2.73)	81.51 (3.66)	81.25 (3.61)
	0.20	68.47 (4.79)	87.61 (0.99)	86.21 (1.53)	89.04 (0.97)	88.93 (1.0)	40.94 (4.75)	78.25 (1.53)	**89.49** (1.09)	89.41 (1.1)
	0.30	82.17 (2.32)	87.62 (0.8)	88.0 (1.2)	89.81 (0.87)	89.74 (0.89)	44.89 (5.34)	82.11 (0.81)	**91.83** (0.86)	91.79 (0.83)
	0.50	88.18 (1.0)	87.84 (0.96)	89.04 (1.0)	89.98 (1.0)	89.94 (0.99)	37.33 (2.51)	85.67 (0.98)	**93.79** (0.91)	93.77 (0.9)
	0.80	89.65 (1.49)	87.88 (1.4)	89.68 (1.45)	89.97 (1.42)	89.97 (1.41)	33.93 (1.16)	88.34 (1.39)	94.91 (1.01)	**94.93** (1.0)
3	0.02	12.2 (4.75)	**79.0** (2.75)	67.07 (6.07)	78.29 (3.14)	77.83 (3.23)	30.17 (11.33)	60.0 (4.17)	26.92 (9.6)	26.12 (9.88)
	0.05	20.42 (7.03)	**84.61** (1.72)	69.2 (4.38)	83.11 (2.08)	82.5 (2.19)	32.0 (5.96)	66.19 (3.73)	64.75 (6.55)	64.28 (6.54)
	0.10	41.62 (6.59)	86.73 (1.36)	80.73 (3.07)	**87.67** (1.45)	87.45 (1.51)	31.95 (5.56)	71.98 (2.73)	81.97 (3.53)	81.79 (3.5)
	0.20	68.47 (4.79)	87.61 (0.99)	86.21 (1.53)	89.04 (0.97)	88.93 (1.0)	40.94 (4.75)	78.25 (1.53)	**89.52** (1.07)	89.44 (1.1)
	0.30	82.17 (2.32)	87.62 (0.8)	88.0 (1.2)	89.81 (0.87)	89.74 (0.89)	44.89 (5.34)	82.11 (0.81)	**91.8** (0.86)	91.78 (0.84)
	0.50	88.18 (1.0)	87.84 (0.96)	89.04 (1.0)	89.98 (1.0)	89.94 (0.99)	37.33 (2.51)	85.67 (0.98)	**93.78** (0.91)	93.77 (0.92)
	0.80	89.65 (1.49)	87.88 (1.4)	89.68 (1.45)	89.97 (1.42)	89.97 (1.41)	33.93 (1.16)	88.34 (1.39)	94.91 (1.0)	**94.94** (0.98)
4	0.02	12.2 (4.75)	**79.0** (2.75)	67.07 (6.07)	78.29 (3.14)	77.83 (3.23)	30.17 (11.33)	60.0 (4.17)	27.69 (9.72)	26.92 (9.95)
	0.05	20.42 (7.03)	**84.61** (1.72)	69.2 (4.38)	83.11 (2.08)	82.5 (2.19)	32.0 (5.96)	66.19 (3.73)	65.51 (6.41)	65.15 (6.38)
	0.10	41.62 (6.59)	86.73 (1.36)	80.73 (3.07)	**87.67** (1.45)	87.45 (1.51)	31.95 (5.56)	71.98 (2.73)	82.13 (3.47)	81.95 (3.45)
	0.20	68.47 (4.79)	87.61 (0.99)	86.21 (1.53)	89.04 (0.97)	88.93 (1.0)	40.94 (4.75)	78.25 (1.53)	**89.52** (1.08)	89.46 (1.1)
	0.30	82.17 (2.32)	87.62 (0.8)	88.0 (1.2)	89.81 (0.87)	89.74 (0.89)	44.89 (5.34)	82.11 (0.81)	**91.79** (0.86)	91.77 (0.83)
	0.50	88.18 (1.0)	87.84 (0.96)	89.04 (1.0)	89.98 (1.0)	89.94 (0.99)	37.33 (2.51)	85.67 (0.98)	**93.78** (0.92)	**93.78** (0.92)
	0.80	89.65 (1.49)	87.88 (1.4)	89.68 (1.45)	89.97 (1.42)	89.97 (1.41)	33.93 (1.16)	88.34 (1.39)	94.92 (1.0)	**94.94** (1.0)

Table 13: Accuracy of various SSL methods on the Salinas A dataset. We pick $k^{(1)} = 50$, $k^{(2)} = 30$, $k^{(3)} = 20$ and $k^{(4)} = 10$. Proposed methods are in bold.

q	rate	Laplace	Poisson	**IP-QC**	**CP-QC**	**IP-SC**	**CP-SC**	**IP-CC**	**CP-CC**
2	1	58.08 (8.37)	57.12 (7.32)	**60.63** (7.49)	58.79 (7.96)	59.53 (7.92)	58.65 (8.04)	58.82 (8.11)	58.5 (8.14)
	2	66.85 (5.49)	57.32 (6.44)	**67.72** (5.37)	67.26 (5.44)	67.23 (5.57)	67.21 (5.47)	66.95 (5.65)	67.11 (5.47)
	5	73.46 (2.31)	56.83 (5.31)	**73.7** (2.34)	73.62 (2.32)	73.66 (2.4)	73.59 (2.28)	73.62 (2.45)	73.55 (2.24)
	10	75.86 (1.82)	56.08 (5.31)	76.2 (1.82)	76.1 (1.82)	**76.24** (1.82)	76.05 (1.83)	76.18 (1.81)	75.98 (1.82)
	20	77.61 (1.15)	56.2 (4.25)	**78.55** (1.38)	77.95 (1.15)	78.35 (1.33)	77.85 (1.15)	78.2 (1.19)	77.77 (1.14)
	50	79.6 (0.88)	56.44 (3.93)	**80.91** (0.89)	80.08 (0.91)	80.72 (0.92)	79.96 (0.89)	80.48 (0.93)	79.85 (0.9)
	100	80.86 (0.57)	56.06 (2.98)	**82.35** (0.63)	81.46 (0.54)	82.14 (0.61)	81.32 (0.54)	81.9 (0.61)	81.17 (0.55)
3	1	58.08 (8.37)	57.12 (7.32)	**60.98** (7.28)	59.25 (7.54)	59.73 (7.89)	59.0 (7.85)	58.81 (8.09)	58.67 (8.07)
	2	66.85 (5.49)	57.32 (6.44)	**67.75** (5.42)	67.45 (5.44)	67.26 (5.6)	67.32 (5.46)	66.85 (5.77)	67.22 (5.46)
	5	73.46 (2.31)	56.83 (5.31)	73.59 (2.36)	**73.65** (2.35)	73.61 (2.42)	73.63 (2.34)	73.58 (2.48)	73.59 (2.27)
	10	75.86 (1.82)	56.08 (5.31)	76.09 (1.88)	**76.21** (1.81)	76.15 (1.84)	76.14 (1.83)	76.15 (1.84)	76.06 (1.83)
	20	77.61 (1.15)	56.2 (4.25)	**78.52** (1.51)	78.14 (1.18)	78.42 (1.43)	78.02 (1.17)	78.26 (1.31)	77.87 (1.15)
	50	79.6 (0.88)	56.44 (3.93)	**80.95** (0.91)	80.37 (0.89)	80.83 (0.94)	80.18 (0.9)	80.64 (0.93)	80.0 (0.9)
	100	80.86 (0.57)	56.06 (2.98)	**82.47** (0.7)	81.82 (0.56)	82.33 (0.62)	81.61 (0.56)	82.1 (0.61)	81.35 (0.55)
4	1	58.08 (8.37)	57.12 (7.32)	**60.88** (7.37)	59.7 (7.13)	59.7 (7.93)	59.12 (7.71)	58.78 (8.12)	58.7 (8.03)
	2	66.85 (5.49)	57.32 (6.44)	**67.7** (5.43)	67.55 (5.47)	67.21 (5.65)	67.35 (5.49)	66.79 (5.8)	67.24 (5.49)
	5	73.46 (2.31)	56.83 (5.31)	73.58 (2.36)	**73.65** (2.37)	73.6 (2.42)	73.62 (2.36)	73.57 (2.48)	73.6 (2.26)
	10	75.86 (1.82)	56.08 (5.31)	76.07 (1.89)	**76.31** (1.8)	76.15 (1.85)	76.18 (1.81)	76.14 (1.85)	76.1 (1.83)
	20	77.61 (1.15)	56.2 (4.25)	**78.52** (1.51)	78.23 (1.18)	78.42 (1.44)	78.09 (1.18)	78.26 (1.33)	77.91 (1.15)
	50	79.6 (0.88)	56.44 (3.93)	**80.95** (0.93)	80.53 (0.87)	80.84 (0.93)	80.31 (0.91)	80.65 (0.94)	80.06 (0.91)
	100	80.86 (0.57)	56.06 (2.98)	**82.45** (0.7)	82.03 (0.59)	82.34 (0.63)	81.77 (0.58)	82.11 (0.61)	81.45 (0.54)

Table 14: Accuracy of various SSL methods on the Salinas A dataset. We pick $k^{(1)} = 50$, $k^{(2)} = 30$, $k^{(3)} = 20$ and $k^{(4)} = 10$. Proposed methods are in bold.

j	rate	Laplace	Poisson	WNLL	Properly	p-Lap	RW	CK	**IP-VQC (2)**	**IP-VQC (3)**
1	1	59.28 (8.54)	58.31 (6.46)	**64.13** (6.05)	64.1 (6.04)	60.26 (5.44)	63.1 (5.14)	28.5 (5.98)	59.93 (8.28)	60.38 (7.94)
	2	66.82 (5.35)	56.76 (7.03)	**67.54** (5.04)	67.42 (5.1)	64.65 (5.13)	66.94 (4.76)	33.05 (6.65)	67.04 (5.19)	67.19 (5.18)
	5	73.74 (2.71)	55.56 (5.89)	73.42 (3.07)	73.14 (3.15)	72.26 (3.07)	73.7 (2.6)	46.37 (5.32)	**73.84** (2.81)	73.78 (2.85)
	10	75.88 (1.67)	56.49 (5.18)	75.81 (1.73)	75.32 (1.81)	74.8 (1.85)	75.98 (1.73)	55.54 (4.27)	**76.15** (1.72)	76.12 (1.87)
	20	77.44 (1.37)	55.99 (4.62)	**78.23** (1.4)	77.56 (1.58)	77.51 (1.61)	77.99 (1.22)	66.04 (3.1)	78.1 (1.38)	78.18 (1.35)
	50	79.58 (0.94)	56.69 (4.19)	**80.87** (0.9)	80.21 (0.93)	80.36 (0.88)	79.1 (0.85)	75.21 (1.76)	80.45 (0.96)	80.77 (0.98)
	100	80.96 (0.73)	55.83 (2.75)	82.1 (0.63)	81.88 (0.66)	82.12 (0.61)	79.27 (0.7)	79.82 (1.02)	81.94 (0.71)	**82.34** (0.73)
2	1	59.28 (8.54)	58.31 (6.46)	**64.13** (6.05)	64.1 (6.04)	60.26 (5.44)	63.1 (5.14)	28.5 (5.98)	61.88 (7.11)	62.23 (6.78)
	2	66.82 (5.35)	56.76 (7.03)	67.54 (5.04)	67.42 (5.1)	64.65 (5.13)	66.94 (4.76)	33.05 (6.65)	67.53 (5.07)	**67.68** (5.12)
	5	73.74 (2.71)	55.56 (5.89)	73.42 (3.07)	73.14 (3.15)	72.26 (3.07)	73.7 (2.6)	46.37 (5.32)	**73.94** (2.84)	73.86 (2.85)
	10	75.88 (1.67)	56.49 (5.18)	75.81 (1.73)	75.32 (1.81)	74.8 (1.85)	75.98 (1.73)	55.54 (4.27)	**76.23** (1.76)	76.14 (1.81)
	20	77.44 (1.37)	55.99 (4.62)	78.23 (1.4)	77.56 (1.58)	77.51 (1.61)	77.99 (1.22)	66.04 (3.1)	78.34 (1.31)	**78.4** (1.37)
	50	79.58 (0.94)	56.69 (4.19)	80.87 (0.9)	80.21 (0.93)	80.36 (0.88)	79.1 (0.85)	75.21 (1.76)	80.87 (0.98)	**80.98** (1.01)
	100	80.96 (0.73)	55.83 (2.75)	82.1 (0.63)	81.88 (0.66)	82.12 (0.61)	79.27 (0.7)	79.82 (1.02)	82.41 (0.72)	**82.53** (0.75)

Table 15: Accuracy of various SSL methods on the MNIST dataset. We pick $k^{(\ell)} = 30 - (\ell - 1) \cdot 10$ for $1 \leq \ell \leq 3$. Proposed methods are in bold.

q	rate	Laplace	Poisson	**IP-QC**	**CP-QC**	**IP-SC**	**CP-SC**	**IP-CC**	**CP-CC**
2	0.02	97.07 (0.07)	96.8 (0.06)	**97.33** (0.07)	97.18 (0.07)	97.24 (0.07)	97.16 (0.07)	97.18 (0.07)	97.14 (0.07)
	0.05	97.37 (0.05)	96.85 (0.04)	**97.62** (0.06)	97.49 (0.05)	97.55 (0.05)	97.46 (0.05)	97.49 (0.05)	97.44 (0.05)
	0.10	97.58 (0.04)	96.85 (0.04)	**97.8** (0.04)	97.69 (0.04)	97.74 (0.04)	97.67 (0.04)	97.69 (0.04)	97.64 (0.04)
	0.20	97.81 (0.04)	96.87 (0.04)	**98.0** (0.04)	97.91 (0.04)	97.95 (0.04)	97.89 (0.04)	97.91 (0.04)	97.87 (0.04)
	0.30	97.92 (0.04)	96.87 (0.05)	**98.1** (0.04)	98.01 (0.05)	98.06 (0.04)	98.0 (0.04)	98.02 (0.04)	97.98 (0.04)
	0.50	98.08 (0.06)	96.87 (0.08)	**98.24** (0.06)	98.16 (0.06)	98.21 (0.06)	98.15 (0.06)	98.17 (0.06)	98.13 (0.06)
	0.80	98.25 (0.09)	96.9 (0.12)	**98.39** (0.09)	98.32 (0.09)	98.36 (0.09)	98.31 (0.09)	98.33 (0.09)	98.29 (0.09)
3	0.02	97.07 (0.07)	96.8 (0.06)	**97.36** (0.07)	97.29 (0.08)	97.26 (0.07)	97.24 (0.07)	97.19 (0.07)	97.19 (0.07)
	0.05	97.37 (0.05)	96.85 (0.04)	**97.64** (0.06)	97.59 (0.06)	97.56 (0.05)	97.54 (0.06)	97.5 (0.05)	97.48 (0.06)
	0.10	97.58 (0.04)	96.85 (0.04)	**97.82** (0.04)	97.77 (0.04)	97.76 (0.04)	97.74 (0.04)	97.7 (0.04)	97.69 (0.04)
	0.20	97.81 (0.04)	96.87 (0.04)	**98.01** (0.04)	97.98 (0.04)	97.97 (0.04)	97.95 (0.04)	97.92 (0.04)	97.91 (0.04)
	0.30	97.92 (0.04)	96.87 (0.05)	**98.1** (0.04)	98.07 (0.04)	98.07 (0.04)	98.05 (0.04)	98.02 (0.04)	98.02 (0.05)
	0.50	98.08 (0.06)	96.87 (0.08)	**98.24** (0.06)	98.21 (0.06)	98.21 (0.06)	98.19 (0.06)	98.18 (0.06)	98.17 (0.06)
	0.80	98.25 (0.09)	96.9 (0.12)	**98.38** (0.09)	98.36 (0.1)	98.37 (0.09)	98.34 (0.09)	98.34 (0.09)	98.32 (0.09)

Table 16: Accuracy of various SSL methods on the MNIST dataset. We pick $k^{(\ell)} = 30 - (\ell - 1) \cdot 10$ for $1 \leq \ell \leq 3$. Proposed methods are in bold.

j	rate	Laplace	Poisson	WNLL	Properly	p-Lap	RW	CK	**IP-VQC (2)**	**IP-VQC** (3)
1	0.02	97.06 (0.09)	96.79 (0.07)	96.55 (0.09)	94.76 (0.17)	94.48 (0.17)	97.15 (0.1)	95.34 (0.16)	97.17 (0.09)	**97.2** (0.09)
	0.05	97.37 (0.06)	96.85 (0.05)	97.2 (0.05)	94.49 (0.12)	95.49 (0.1)	97.37 (0.07)	96.46 (0.08)	97.49 (0.05)	**97.52** (0.05)
	0.10	97.59 (0.04)	96.86 (0.04)	97.58 (0.05)	95.59 (0.08)	96.88 (0.06)	97.45 (0.05)	97.18 (0.06)	97.69 (0.04)	**97.73** (0.04)
	0.20	97.8 (0.04)	96.87 (0.04)	97.86 (0.04)	97.08 (0.05)	97.71 (0.04)	97.5 (0.05)	97.68 (0.04)	97.9 (0.04)	**97.93** (0.04)
	0.30	97.92 (0.05)	96.87 (0.05)	97.98 (0.05)	97.61 (0.06)	97.88 (0.05)	97.51 (0.05)	97.88 (0.05)	98.02 (0.05)	**98.04** (0.05)
	0.50	98.08 (0.06)	96.86 (0.06)	98.11 (0.06)	98.01 (0.06)	98.07 (0.06)	97.51 (0.06)	98.09 (0.06)	98.17 (0.06)	**98.19** (0.06)
	0.80	98.22 (0.1)	96.87 (0.14)	98.23 (0.1)	98.22 (0.11)	98.23 (0.1)	97.52 (0.13)	98.24 (0.11)	98.31 (0.11)	**98.33** (0.11)
2	0.02	97.06 (0.09)	96.79 (0.07)	96.55 (0.09)	94.76 (0.17)	94.48 (0.17)	97.15 (0.1)	95.34 (0.16)	97.31 (0.09)	**97.34** (0.09)
	0.05	97.37 (0.06)	96.85 (0.05)	97.2 (0.05)	94.49 (0.12)	95.49 (0.1)	97.37 (0.07)	96.46 (0.08)	97.62 (0.05)	**97.64** (0.05)
	0.10	97.59 (0.04)	96.86 (0.04)	97.58 (0.05)	95.59 (0.08)	96.88 (0.06)	97.45 (0.05)	97.18 (0.06)	97.8 (0.04)	**97.82** (0.04)
	0.20	97.8 (0.04)	96.87 (0.04)	97.86 (0.04)	97.08 (0.05)	97.71 (0.04)	97.5 (0.05)	97.68 (0.04)	97.99 (0.04)	**98.0** (0.04)
	0.30	97.92 (0.05)	96.87 (0.05)	97.98 (0.05)	97.61 (0.06)	97.88 (0.05)	97.51 (0.05)	97.88 (0.05)	**98.1** (0.05)	**98.1** (0.05)
	0.50	98.08 (0.06)	96.86 (0.06)	98.11 (0.06)	98.01 (0.06)	98.07 (0.06)	97.51 (0.06)	98.09 (0.06)	**98.24** (0.06)	**98.24** (0.05)
	0.80	98.22 (0.1)	96.87 (0.14)	98.23 (0.1)	98.22 (0.11)	98.23 (0.1)	97.52 (0.13)	98.24 (0.11)	**98.37** (0.11)	**98.37** (0.11)