

# Higher-Order Regularization Learning on Hypergraphs

Adrien Weihs<sup>\*1</sup>, Andrea L. Bertozzi<sup>1</sup>, and Matthew Thorpe<sup>2</sup>

<sup>1</sup>Department of Mathematics,  
University of California Los Angeles,  
Los Angeles, CA 90095, USA.

<sup>2</sup>Department of Statistics,  
University of Warwick,  
Coventry, CV4 7AL, UK.

November 2025

## Abstract

Higher-Order Hypergraph Learning (HOHL) was recently introduced as a principled alternative to classical hypergraph regularization, enforcing higher-order smoothness via powers of multiscale Laplacians induced by the hypergraph structure. Prior work established the well- and ill-posedness of HOHL through an asymptotic consistency analysis in geometric settings. We extend this theoretical foundation by proving the consistency of a truncated version of HOHL and deriving explicit convergence rates when HOHL is used as a regularizer in fully supervised learning. We further demonstrate its strong empirical performance in active learning and in datasets lacking an underlying geometric structure, highlighting HOHL's versatility and robustness across diverse learning settings.

**Keywords and phrases.** Hypergraph learning, Semi-supervised learning, Sobolev regularization, Asymptotic consistency, Multiscale learning, Non-Euclidean data, Active learning, Graph Laplacians

**Mathematics Subject Classification.** 49J55, 49J45, 62G20, 65N12

---

<sup>\*</sup>Corresponding author. Email: [weihs@math.ucla.edu](mailto:weihs@math.ucla.edu)

# 1 Introduction

Graphs play a foundational role in machine learning, enabling effective modeling of relational data across a range of tasks—from semi-supervised learning and clustering to recommendation systems and manifold learning, e.g. [11, 27, 50, 57, 59, 90, 96–98]. However, many real-world phenomena involve more complex interactions among sets of nodes, that are not fully captured by pairwise edges. Hypergraphs extend graphs by allowing hyperedges to connect arbitrary subsets of nodes, and hypergraph-based methods are used broadly in various areas of science such as in [17, 25, 30, 31, 46, 56, 62, 67, 69, 73, 92, 94, 95].

A central research question concerns the comparison between hypergraph and graph-based learning methods. Many such comparisons are grounded in discrete arguments (e.g., [1, 18, 48, 49, 61]). More recently, asymptotic consistency frameworks—a popular technique for analyzing graph-based methods by relating discrete energies to continuum variational limits (e.g., [4, 9, 19, 23, 41, 75, 77, 80, 89])—have been extended to the hypergraph regularization setting [73, 87]. This continuum perspective allows for a principled assessment of the role of hypergraph structures, supports a classification of hypergraph learning algorithms [87, Figure 2], and enables a clearer understanding of the regularization behavior underlying complex discrete formulations.

In the analysis of graph- and hypergraph-based regularization, it is useful to distinguish between two complementary components of a regularizer: (1) the support of interactions, i.e., which nodes influence one another (determined by the graph or hypergraph topology), and (2) the interaction mechanism, i.e., how these influences are aggregated or penalized (e.g., via first-order differences, higher-order derivatives, or more general nonlinear terms). Classical hypergraph learning methods typically enrich the interaction support—by allowing edges to connect sets of nodes rather than pairs—but still rely on first-order, pairwise-like regularization mechanisms [87, 95].

In this context, Higher-Order Hypergraph Learning (HOHL) was introduced as a method that more effectively leverages hypergraph structure—not only by modifying which interactions are considered, but also by altering the nature of such interactions. Specifically, HOHL decomposes the hypergraph into a sequence of subgraphs that capture interactions at multiple scales. On each subgraph, a distinct regularization strength is applied, allowing the model to enforce higher-order smoothness in a structured and scale-aware manner. In doing so, HOHL exploits the full expressive potential of the hypergraph more fully and effectively. From an analytical perspective, HOHL is shown to converge to a higher-order Sobolev semi-norm, making it genuinely distinct from other hypergraph methods [73, 95] that asymptotically recover the standard  $W^{1,p}$  regularization.

In this paper, we extend both the theoretical and computational analysis of HOHL. On the theoretical side, we prove that when HOHL is used as a regularizer in the fully supervised learning setting, it yields explicit rates of convergence between the learned function and the ground-truth target. Furthermore, we analyze a truncated version of the HOHL energy—commonly employed in practice due to its reduced computational complexity—and establish that it remains consistent, converging to the same higher-order continuum limit as the full model.

On the computational side, we demonstrate that HOHL preserves the quadratic form characteristic of Laplace learning and can, in fact, be interpreted as Laplace learning on a specially constructed graph. This equivalence implies that all existing computational techniques developed for Laplace learning are directly applicable to HOHL, enabling seamless integration into established workflows. In particular, we highlight this drop-in compatibility through an active learning application, where HOHL strongly outperforms traditional Laplacian-based approaches. Finally, we extend the HOHL framework to settings where the hypergraph is not embedded in some underlying metric space. This generalization necessitates a shift in the notion of scale-aware regularization, but continues to yield strong performance, achieving state-of-the-art results on several standard hypergraph benchmarks.

## 1.1 Contributions

Our main contributions are as follows:

1. **Theoretical Guarantees for Supervised Learning:** We prove that using HOHL as a regularizer in the fully supervised setting yields explicit convergence rates between the learned function and the ground-truth target.

2. **Consistency of Truncated HOHL:** We analyze a truncated version of HOHL, commonly used in practice for its computational efficiency, and establish that it remains consistent with the full model by converging to the same higher-order continuum limit.
3. **Connection to Laplace Learning:** We show that HOHL preserves the quadratic form of Laplace learning and can be interpreted as Laplace learning on a specially constructed graph, making all standard computational techniques for Laplace learning directly applicable.
4. **Plug-and-Play Use in Active Learning:** We demonstrate that HOHL can serve as a drop-in replacement for Laplace learning in existing pipelines, highlighting its advantages through strong empirical performance in active learning tasks.
5. **Extension Beyond Geometric Hypergraphs:** We generalize HOHL to hypergraphs without an underlying metric structure by redefining the notion of multiscale regularization, achieving state-of-the-art results on standard hypergraph learning benchmarks.

## 1.2 Related works

A growing body of work has focused on the asymptotic consistency and continuum analysis of graph-based regularization in the large-sample regime. These efforts include convergence results for total variation on graphs [36], graph cuts and Cheeger-type problems [34, 38, 39, 64], the Mumford–Shah functional [15], and empirical risk minimization [33]. In the semi-supervised setting, particular attention has been given to  $p$ -Laplace learning [77], fractional Laplacian methods [89], Lipschitz learning [8, 10, 53, 68], game-theoretic formulations [9], Poisson learning [7, 11], reweighted Laplacians [74], and truncated energy models [2, 3]. These developments reflect a general trend toward understanding the behavior of discrete algorithms through the lens of continuum variational principles. Recently, such analyses have been extended to the hypergraph setting [73, 87]

Consistency between discrete energies  $\mathcal{E}_n$ , defined for functions  $v_n : \Omega_n \rightarrow \mathbb{R}$ , and a corresponding continuum energy  $\mathcal{E}_\infty$ , defined on functions  $v : \Omega \rightarrow \mathbb{R}$ , can be established through several analytical approaches:

- *Pointwise convergence* [4, 19, 41, 44, 45, 75, 80] examines whether  $\mathcal{E}_n(v|_{\Omega_n}) \rightarrow \mathcal{E}_\infty(v)$  as  $n \rightarrow \infty$ , for sufficiently smooth functions  $v : \Omega \rightarrow \mathbb{R}$ . A related approach considers the pointwise convergence of the associated Euler–Lagrange operators [88].
- *Spectral convergence* [4, 12, 32, 65, 76, 85, 86] analyzes the convergence of the eigenvalues and eigenfunctions of the discrete operator associated with (through Euler–Lagrange equations)  $\mathcal{E}_n$  to those of the limiting operator appearing in  $\mathcal{E}_\infty$ .
- *Variational convergence* [9, 20, 23, 36–39, 77, 79, 82] concerns the convergence of minimizers of  $\mathcal{E}_n$  to those of  $\mathcal{E}_\infty$ , typically formalized through  $\Gamma$ -convergence [6]. Among the three notions, it is often the most relevant in semi-supervised learning, where the final label assignments are derived from minimizers of the objective functional.

In this work, we focus on the latter two modes of convergence. In particular, to establish the variational convergence of our truncated energies, we analyze the spectral properties of the HOHL Laplacian [87]. Our results in the fully supervised setting are also of variational type, providing convergence guarantees for minimizers of the discrete energies.

While much of the literature has focused on consistency, in the graph-based setting, recent works established convergence rates in terms of various parameters such as the number of points  $n$ , the labeling rate, the graph connectivity parameter  $\varepsilon$  and the smoothness of the target function [14, 24, 88]. These rates offer important theoretical guarantees for practical applications, where the dataset is finite and the discrete approximation error must be controlled. In this work, we extend such results to the HOHL framework, similarly to [35], showing explicit convergence rates between the discrete minimizers and the continuum ground truth under suitable regularity assumptions.

Beyond rates, computational efficiency is a key concern for applications. In practice, Laplace learning and related graph-based methods often rely on a spectrally truncated energy formulation. While these truncations are computationally efficient and widely adopted in large-scale settings, their theoretical justification

has largely remained heuristic [2, 5, 60]. In this work, we contribute to closing this gap by showing that even when the HOHL energy is truncated, it remains variationally consistent with the full model and converges to the same continuum limit.

Our final computational result establishes a connection between the HOHL Laplacian, and a broad body of work on graph reweighting [13, 74] (in classical models) and graph rewiring [26, 42, 51, 52, 63, 81] (in graph neural networks), both of which aim to improve learning performance by structurally modifying the graph. These modifications are often employed to address limitations such as oversmoothing or oversquashing [42]. In the spirit of [1], which advocates for representing hypergraph structure within enriched graph formulations, we show that HOHL can be interpreted as Laplace learning on a modified graph constructed directly from the original hypergraph structure.

## 2 Background

This section presents the mathematical tools used throughout the paper. We begin by recalling the  $TL^p$  space, which provides a natural topology for comparing functions defined on discrete empirical measures to functions on the continuum. We then review key concepts from  $\Gamma$ -convergence theory, which we rely on to study the asymptotic behavior of our discrete variational problems. References for the material presented here include [6, 36, 77, 89].

### 2.1 The $TL^p$ Topology

Let  $\mathcal{P}_p(\Omega)$  denote the set of Borel probability measures on a bounded domain  $\Omega \subset \mathbb{R}^d$  with finite  $p$ -th moment. For each  $\mu \in \mathcal{P}_p(\Omega)$ , we denote by  $L^p(\mu)$  the space of  $\mu$ -measurable functions with finite  $L^p$  norm. A key operation when comparing measures is the pushforward. Given a measurable map  $T : \Omega \rightarrow \mathcal{Z}$  and a measure  $\mu \in \mathcal{P}(\Omega)$ , the pushforward measure  $T_{\#}\mu \in \mathcal{P}(\mathcal{Z})$  is defined by:

$$T_{\#}\mu(A) := \mu(T^{-1}(A)) \quad \text{for all measurable sets } A \subset \mathcal{Z}.$$

**Definition 2.1.** For an underlying domain  $\Omega$ , define the set

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

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

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

where  $\Pi(\mu, \nu)$  is the set of couplings between  $\mu$  and  $\nu$ .

This framework allows us to treat discrete functions—defined on sampled data—as elements of a well-defined metric space and to compare them to their continuum counterparts in a stable way. The topology is closely related to the  $p$ -Wasserstein distance [70, 83] on the graph of the function.

A useful characterization of convergence in  $TL^p$  is the following [36, Proposition 3.12].

**Proposition 2.2.** Let  $(\mu_n, u_n) \in TL^p$  be a sequence and  $(\mu, u) \in TL^p$ . Assume that  $\mu$  is absolutely continuous with respect to the Lebesgue measure. Then the following are equivalent:

1.  $(\mu_n, u_n) \rightarrow (\mu, u)$  in  $TL^p$ ;
2.  $\mu_n$  converges weakly to  $\mu$  and there exists a sequence of transport maps  $\{T_n\}_{n=1}^{\infty}$  with  $(T_n)_{\#}\mu = \mu_n$  and  $\int_{\Omega} |x - T_n(x)| dx \rightarrow 0$  such that

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

To apply this result, we rely on the following result (see [32, Theorem 2] based on [40]) establishing that such transport maps exist for empirical measures constructed from i.i.d. samples.

**Theorem 2.3** (Existence of transport maps). *Assume that  $\Omega$  is the unit torus  $\mathbb{R}^d/\mathbb{Z}^d$ ,  $x_i \stackrel{\text{iid}}{\sim} \mu \in \mathcal{P}(\Omega)$  where  $\mu$  has a density that is bounded above and below by positive constants. 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  $\mu$  to  $\mu_n$  such that:*

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

The assumptions required in the above theorem correspond to conditions **S.1**, **M.1**, **M.2**, and **D.1** introduced later in the paper. Taken together, these results enable a rigorous comparison between discrete functionals defined over sample-based measures and their continuum limits.

## 2.2 $\Gamma$ -Convergence of Functionals

To analyze the asymptotic behavior of our variational formulations, we use  $\Gamma$ -convergence, a notion from the calculus of variations that captures the convergence of minimization problems.

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

1. *For every  $z \in Z$  and every sequence  $\{z_n\}$  with  $d_Z(z_n, z) \rightarrow 0$ :*

$$\liminf_{n \rightarrow \infty} F_n(z_n) \geq F(z);$$

2. *For every  $z \in Z$ , there exists a sequence  $\{z_n\}$  with  $d_Z(z_n, z) \rightarrow 0$  and*

$$\limsup_{n \rightarrow \infty} F_n(z_n) \leq F(z).$$

This notion of convergence ensures that the minimizers of  $F_n$  converge (in a suitable sense) to minimizers of  $F$ , provided a compactness condition holds.

**Definition 2.5.** *We say that a sequence of functionals  $F_n : Z \rightarrow \mathbb{R}$  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  $\infty$ . Suppose that the functionals satisfy the compactness property and that they  $\Gamma$ -converge to  $F : Z \mapsto [0, \infty]$ . Then*

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

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

$$(1) \quad \lim_{n \rightarrow \infty} \left( F_n(z_n) - \inf_{z \in Z} F_n(z) \right) = 0$$

*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 (1) converges to the unique minimizer of  $F$ .*

In this work, we show that our discrete energies  $\Gamma$ -converge to continuum energies in the  $\text{TL}^p$ -topology. This forms the backbone of our theoretical analysis, allowing us to rigorously link discrete regularization schemes to their continuum analogues.

Lastly, the following result shows that  $\Gamma$ -convergence is stable with respect to continuous perturbations.

**Proposition 2.7** (Convergence of minimizers). *Suppose that  $F_n : Z \mapsto [0, \infty]$   $\Gamma$ -converge to  $F : Z \mapsto [0, \infty]$ . Furthermore, assume that  $G : Z \mapsto [0, \infty]$  is continuous. Then,  $F_n + G$   $\Gamma$ -converges to  $F + G$ .*

### 3 Main results

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

#### 3.1 Hypergraphs

A hypergraph  $G$  is a pair  $G = (V, E)$ , where  $V$  denotes the set of vertices and  $E$  is a collection of subsets  $e \subseteq V$ , called hyperedges. We say that all vertices within the same hyperedge  $e$  are connected and denote the weight of hyperedge  $e$  by  $w_0(e) \geq 0$  and its degree/size by  $|e|$ . We write  $V = \{v_i\}_{i=1}^{|V|}$ .

A special case of hypergraphs is when  $|e| = 2$  for all  $e \in E$ . In this case,  $(V, E)$  is called a graph and every  $e$  represents a pairwise relationship between vertices (see Figure 1). Graphs can also be weighted and we usually use the representation  $G = (V, W)$  where  $W \in \mathbb{R}^{|V| \times |V|}$  is a symmetric matrix with entries  $w_{ij} = w_0(e)$  if  $e = \{v_i, v_j\}$ . On graphs, we define the (unnormalized) Laplacian  $L$  as

$$L = D - W$$

where  $D$  is the diagonal matrix with entries  $d_{ii} = \sum_{j=1}^{|V|} w_{ij}$ .

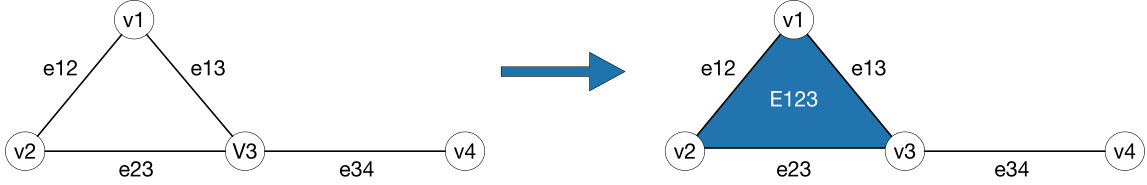


Figure 1: From graphs to hypergraphs (from [87]). Left: In the graph, the vertices  $v_1$ ,  $v_2$ , and  $v_3$  are all connected pairwise. Right: A single hyperedge is added connecting all three vertices, transitioning from a graph to a hypergraph representation.

We now introduce the hypergraph-to-graph deconstruction that is the foundation of HOHL. Let  $(V, E)$  be a hypergraph and define  $q = \max_{e \in E} |e| - 1$  as the maximum hyperedge size minus one. For each  $k \in \{1, \dots, q\}$ , we construct a corresponding skeleton graph  $G^{(k)} = (V, E^{(k)})$  with

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

that is,  $G^{(k)}$  contains all pairwise edges induced by hyperedges of size  $k + 1$ . We refer to Figure 2 for a visual representation of the decomposition. Let  $L^{(k)}$  denote the graph Laplacian associated with  $G^{(k)}$ .

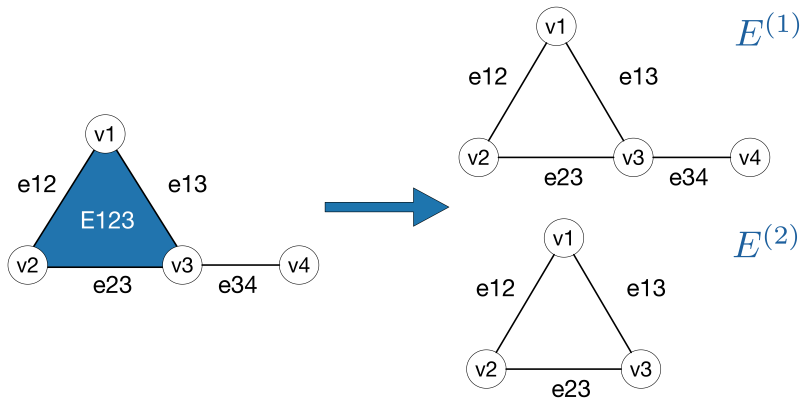


Figure 2: Skeleton graphs with  $q = 2$  (from [87]).

### 3.2 HOHL

On graphs, a widely used regularizer is constructed using the graph Laplacian [84, 98]. For a function  $u : V \rightarrow \mathbb{R}$  (which we also identify with a vector in  $\mathbb{R}^{|V|}$ ), its first-order smoothness is quantified by

$$u^\top L u = \frac{1}{2} \sum_{i,j=1}^{|V|} w_{ij} (u(v_i) - u(v_j))^2.$$

Minimizing this expression encourages  $u$  to take similar values on adjacent vertices. On certain graphs, this functional can be interpreted as a discrete analogue of the Sobolev  $W^{1,2}$  semi-norm, which formalizes the idea of penalizing the first derivative of a function defined on the graph [77]. More generally, the regularizer  $v^\top L^s v$ , with  $s \in \mathbb{R}$ , corresponds to a discrete Sobolev  $W^{s,2}$  semi-norm and penalizes variations of  $v$  up to order  $s$  [23, 89].

The HOHL energy, introduced in [87], extends graph Laplacian regularization to the hypergraph setting. It is defined as

$$(2) \quad u^\top \left[ \sum_{k=1}^q \lambda_k (L^{(k)})^k \right] u =: u^\top \mathcal{L}_{\text{dis}}^{(q)} u,$$

for  $u \in \mathbb{R}^n$ , where  $0 < p_1 < \dots < p_q$  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  $\{p_k\}_{k=1}^q$ . This energy imposes a hierarchical, scale-aware regularization: for each skeleton graph  $G^{(k)}$ , the corresponding Laplacian power  $(L^{(k)})^{p_k}$  enforces smoothness at a specific scale, with the index  $k$  controlling the granularity of the regularization.

We now discuss the geometric setting, where  $V \subset \mathbb{R}^d$ , and the hyperedge set  $E$  is not given a priori. In such cases, it is common to construct  $E$  using geometric principles. The underlying intuition is that a meaningful hyperedge should connect vertices that are close in some metric space.

In the graph setting, this idea is typically implemented via  $k$ -nearest neighbor ( $k$ -NN) graphs [84] or random geometric graphs [66], both of which rely on locality: edges are formed either by linking the  $k$  nearest neighbors or by connecting points within an  $\varepsilon$ -radius neighborhood. Analogous locality-based constructions for hypergraphs have been proposed, e.g., in [73]; see also [31] for a broader discussion. A notable instance is also the random geometric hypergraph model introduced in [87].

As established in [87], the hierarchical, scale-aware regularization principle underlying HOHL admits an effective surrogate in geometric settings via a multiscale graph construction, as proposed in [59]. In what follows, we introduce this alternative formulation.

Let  $\Omega_n = \{x_i\}_{i=1}^n \subset \Omega \subset \mathbb{R}^d$  be a set of  $n$  feature vectors, where we assume that  $x_i \stackrel{\text{i.i.d.}}{\sim} \mu \in \mathcal{P}(\Omega)$ . We adopt the same probabilistic framework as in [89]. Specifically, we consider a probability space  $(\Omega, \mathbb{P})$  whose elements are infinite sequences  $\{x_i\}_{i=1}^\infty$ . Our results are stated in terms of the measure  $\mathbb{P}$ , establishing that the desired properties hold on a high-probability subset  $\mathcal{X} \subset \Omega$  consisting of such sequences. For a set  $E$ , we denote its complement by  $E^c$ .

Given a length-scale  $\varepsilon > 0$ , and a kernel function  $\eta$ , we define the edge weights  $w_{\varepsilon,ij}$  between vertices  $x_i$  and  $x_j$  by

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

Let  $D_{n,\varepsilon}$  be the diagonal degree matrix with entries  $d_{n,\varepsilon,ii} = \sum_{j=1}^n w_{\varepsilon,ij}$ , and define the normalizing constant

$$\sigma_\eta = \frac{1}{d} \int_{\mathbb{R}^d} \eta(|h|) |h|^2 dh < \infty.$$

The (unnormalized) graph Laplacian is then given by

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

We note that this is the rescaled version of  $L$ , i.e.  $\Delta_{n,\varepsilon} = \frac{2}{\sigma_{\eta n \varepsilon^{d+2}}} L$ . With an abuse of notation, this Laplacian can be interpreted either as a matrix  $\Delta_{n,\varepsilon} \in \mathbb{R}^{n \times n}$  or as an operator  $\Delta_{n,\varepsilon} : L^2(\mu_n) \rightarrow L^2(\mu_n)$ , where  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$  is the empirical measure.

For functions  $u_n, v_n : \Omega_n \rightarrow \mathbb{R}$ , we define the  $L^2(\mu_n)$  inner product by

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

Such functions can be regarded as vectors in  $\mathbb{R}^n$ ; in what follows, we will use the notation  $u_n$  both for the function  $u_n : \Omega_n \rightarrow \mathbb{R}$  and for the associated vector in  $\mathbb{R}^n$ .

We denote by  $\{(a_{n,\varepsilon,k}, \phi_{n,\varepsilon,k})\}_{k=1}^n$  the eigenpairs of  $\Delta_{n,\varepsilon}$ , where the eigenvalues are ordered nondecreasingly:  $0 = a_{n,\varepsilon,1} < a_{n,\varepsilon,2} \leq a_{n,\varepsilon,3} \leq \dots \leq a_{n,\varepsilon,n}$  (with strict inequality between  $a_{n,\varepsilon,1}$  and  $a_{n,\varepsilon,2}$  whenever the graph  $(\Omega_n, W_{n,\varepsilon})$  is connected). The corresponding eigenfunctions  $\{\phi_{n,\varepsilon,k}\}_{k=1}^n$  form an orthonormal basis of  $L^2(\mu_n)$ .

Given the Laplacians defined above, the surrogate for HOHL (2) is

$$(3) \quad v^\top \left[ \sum_{k=1}^q \lambda_k \Delta_{n,\varepsilon^{(k)}}^{p_k} \right] v,$$

where  $\varepsilon^{(1)} > \dots > \varepsilon^{(q)}$ , and  $p_k > 0$  controls the regularity imposed at each scale. We will allow the length-scales to vary with the number of data point, i.e.  $\varepsilon^{(k)} = \varepsilon_n^{(k)}$ , and in this case, we write  $E_n := \{\varepsilon_n^{(k)}\}_{k=1}^q$ . The well and ill-posedness of (3) in semi-supervised learning is precisely characterized in [87, Theorem 3.5] as a function of  $\varepsilon_n^{(q)}$ .

We now define the continuum analogues of our discrete Laplacian operators. Let  $\Delta_\rho$  be the continuum weighted Laplacian operator defined by

$$\Delta_\rho u(x) = -\frac{1}{\rho(x)} \operatorname{div}(\rho^2 \nabla u)(x), \quad x \in \Omega$$

and let  $\{(\beta_i, \psi_i)\}_{i=1}^\infty$  be its associated eigenpairs where  $\beta_1 = 0 < \beta_2 \leq \beta_3 \leq \dots$ . Here  $\rho$  denotes the density of  $\mu$  with respect to Lebesgue measure. We note that  $\{\psi_i\}_{i=1}^\infty$  form a basis of  $L^2(\mu)$  and we also define

$$(4) \quad \mathcal{H}^s(\Omega) = \left\{ h \in L^2(\mu) \mid \|h\|_{\mathcal{H}^s(\Omega)}^2 := \sum_{i=1}^\infty \beta_i^s \langle h, \psi_i \rangle_{L^2(\mu)}^2 < +\infty \right\}.$$

The space  $\mathcal{H}^s(\Omega)$  is closely related to the Sobolev space  $W^{s,2}(\Omega)$  [23, Lemma 17].

**Fully supervised problem with HOHL regularization.** We now turn our attention to the fully supervised problem, where (3) is used as a regularizer. Specifically, for some sequence of points  $\mathbf{g}_n = \{g_i\}_{i=1}^n$ , parameter  $\tau > 0$  and  $v_n : \Omega_n \mapsto \mathbb{R}$ , we define the fully supervised learning problem

$$\mathcal{R}_{n,\tau}^{(\mathbf{g}_n)}(v_n) = \frac{1}{n} \sum_{i=1}^n |v_n(x_i) - g_i|^2 + \tau \sum_{k=1}^q \lambda_k \langle v_n, \Delta_{n,\varepsilon_n^{(k)}}^{p_k} v_n \rangle_{L^2(\mu_n)}.$$

For some  $g \in C^0$  and  $v : \Omega \mapsto \mathbb{R}$ , the continuum counterpart to the above is

$$\mathcal{R}_{\infty,\tau}^{(g)}(u) = \int_{\Omega} |v(x) - g(x)|^2 \rho(x) dx + \tau \sum_{k=1}^q \lambda_k \langle v, \Delta_\rho^{p_k} v \rangle_{L^2(\mu)}.$$

We are mainly interested in the case of noisy labels, i.e. when for some  $g \in C^0(\Omega)$ , we have labels  $\mathbf{y}_n = \{y_i\}_{i=1}^n$  where  $y_i = g(x_i) + \xi_i$  and  $\xi_i \in \mathbb{R}$  are independent and identically distributed sub-Gaussian centered noise. Adopting the terminology of [21], for  $u_{n,\tau}^{(\mathbf{y}_n)}$  and  $u_\tau$  the minimizers of  $\mathcal{R}_{n,\tau}^{(\mathbf{y}_n)}$  and  $\mathcal{R}_{\infty,\tau}^{(g)}$  respectively, we regard  $u_{n,\tau}^{(\mathbf{y}_n)}$  as an estimator of  $g$ . We further decompose its error into a variance and bias component, defined as

$$\|u_{n,\tau}^{(\mathbf{y}_n)} - u_\tau\|_{L^2(\mu_n)} \quad \text{and} \quad \|u_\tau - g\|_{L^2(\mu)}$$

respectively. Intuitively, the variance term quantifies the fluctuation caused by finite sampling, while the bias term measures the approximation error introduced by the regularization.



**Truncated HOHL energies.** We also consider the truncated versions of our energies. We define the matrix  $\mathcal{L}_n^{(q)} = \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k}$  and its continuum counterpart  $\mathcal{L}^{(q)} = \sum_{k=1}^q \lambda_k \Delta_\rho^{p_k}$ . In particular,  $\mathcal{L}_n^{(q)}$  is positive semi-definite and we denote its ordered eigenpairs by  $\{(\beta_{n,i}, \psi_{n,i})\}_{i=1}^n$ . Then,

$$\langle v, \mathcal{L}_n^{(q)} v \rangle_{L^2(\mu_n)} = \sum_{i=1}^n \beta_{n,i} \langle v, \psi_{n,i} \rangle_{L^2(\mu_n)}^2$$

and the truncated energy for some threshold  $T \leq n$  is

$$\sum_{i=1}^T \beta_{n,i} \langle v, \psi_{n,i} \rangle_{L^2(\mu_n)}^2.$$

We define the variational problems

$$(\mathcal{S}\mathcal{J})_{n, E_n, \Psi, T}^{(q, P)}((\nu, v)) = \begin{cases} \sum_{i=1}^T \beta_{n,i} \langle v, \psi_{n,i} \rangle_{L^2(\mu_n)}^2 + \Psi((\nu, v)) & \text{if } \nu = \mu_n \text{ and } \langle v, \psi_{n,k} \rangle_{L^2(\mu_n)} = 0 \\ & \text{for all } k > T, \\ +\infty & \text{else,} \end{cases}$$

and

$$(\mathcal{S}\mathcal{J})_{\infty, \Psi}^{(q, P)}((\nu, v)) = \begin{cases} \sum_{i=1}^{\infty} \left( \sum_{k=1}^q \lambda_k \beta_i^{p_k} \right) \langle v, \psi_i \rangle_{L^2(\mu)}^2 + \Psi((\nu, v)) & \text{if } \nu = \mu, \\ +\infty & \text{else,} \end{cases}$$

where  $\Psi : \text{TL}^2(\Omega) \mapsto \mathbb{R}$  is a continuous function acting as data-fidelity term (for example  $\Psi((\nu, v)) = \int_{\Omega} |v(x) - y(x)|^2 d\nu(x)$  where  $y : \Omega \mapsto \mathbb{R}$  is Lipschitz continuous). The minimizers of  $(\mathcal{S}\mathcal{J})_{n, E_n, \Psi, T}^{(q, P)}((\nu, v))$  are spanned by the first  $T$  eigenvectors  $\psi_{n,i}$ .

### 3.3 Assumptions

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

*Assumptions 1.* Assumption on the space.

**S.1** The feature vector space  $\Omega$  is the unit torus  $\mathbb{R}^d / \mathbb{Z}^d$ .

*Assumptions 2.* Assumptions on the measure.

**M.1** The measure  $\mu$  is a probability measure on  $\Omega$ .

**M.2** There is a continuous Lebesgue density  $\rho$  of  $\mu$  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 we make the following assumptions.

*Assumptions 3.* Assumptions on the data.

**D.1** Feature vectors  $\Omega_n = \{x_i\}_{i=1}^n$  are iid samples from a measure  $\mu$  satisfying **M.1**. We denote by  $\mu_n$  the empirical measure associated to our samples.

The weight function  $\eta$  is assumed to satisfy the following assumptions.

*Assumptions 4.* Assumptions on the weight function or kernel.

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

**W.2** The function  $\eta : [0, \infty) \rightarrow [0, \infty)$  satisfies  $\eta(t) > \frac{1}{2}$  for  $t \leq \frac{1}{2}$ ,  $\eta(t) = 0$  for all  $t \geq 1$  and is decreasing.

The assumption that  $\eta$  has compact support reflects the practical constraint in most applications: for computational efficiency, one typically limits the interaction range between vertices in the hypergraph.

### 3.4 Main results

#### 3.4.1 Fully supervised problem with HOHL regularization

We start by establishing the following rates of convergence between the minimizer  $u_{n,\tau}^{(\mathbf{y}_n)}$  of  $\mathcal{R}_{n,\tau}^{(\mathbf{y}_n)}$  and  $g$ . The function  $u_{n,\tau}^{(\mathbf{y}_n)}$  is the best regularized approximation of  $g$  on the graph given the label noise.

**Theorem 3.1** (Rates between discrete minimizers and labelling function). *Assume that S.1, M.1, M.2, D.1 and W.2 hold. Let  $q \geq 1$ ,  $\{\lambda_k\}_{k=1}^q$  be a sequence of positive numbers,  $P = \{p_k\}_{k=1}^q \subseteq \mathbb{N}$  with  $1 \leq p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)} > 0$ . Furthermore, let  $\rho \in C^\infty$  and assume that  $W_{\varepsilon_n^{(k)}, ii} = 0$ . Let  $\xi_i$  be iid, mean zero, sub-Gaussian random variables,  $g \in C^\infty$  and  $\mathbf{y}_n = \{y_i\}_{i=1}^n$  with  $y_i = g(x_i) + \xi_i$ . Then, for all  $\alpha > 1$  and  $\tau_0$ , there exists  $\varepsilon_0 > 0$  and  $C > c > 0$  such that for all  $E_n$  satisfying*

$$\varepsilon_0 \geq \varepsilon_n^{(1)} \geq \dots \geq \varepsilon_n^{(q)} \geq C \left( \frac{\log(n)}{n} \right)^{1/d},$$

and  $0 < \tau < \tau_0$ , the following holds with probability  $1 - Cn^{-\alpha} - Cne^{-cn(\varepsilon_n^{(q)})^{d+4pq}}$ :

$$(5) \quad \|u_{n,\tau}^{(\mathbf{y}_n)} - g|_{\Omega_n}\|_{L^2(\mu_n)} \leq C \left[ \sum_{k=1}^q \lambda_k \frac{(\varepsilon_n^{(1)})^{2p_1}}{(\varepsilon_n^{(k)})^{2p_k}} \left( \frac{\log(n)}{n(\varepsilon_n^{(k)})^d} \right)^{1/2} + \frac{(\varepsilon_n^{(1)})^{2p_1}}{\tau} + \tau \left( 1 + \sum_{k=1}^q \lambda_k \varepsilon_n^{(k)} \right) \right]$$

where  $u_{n,\tau}^{(\mathbf{y}_n)}$  is the minimizer of  $\mathcal{R}_{n,\tau}^{(\mathbf{y}_n)}$ .

We note from Propositions 4.3 and 4.4 that the variance term scales as

$$\sum_{k=1}^q \lambda_k \frac{(\varepsilon_n^{(1)})^{2p_1}}{(\varepsilon_n^{(k)})^{2p_k}} \left( \frac{\log(n)}{n(\varepsilon_n^{(k)})^d} \right)^{1/2} + \frac{(\varepsilon_n^{(1)})^{2p_1}}{\tau} + \tau \sum_{k=1}^q \lambda_k \varepsilon_n^{(k)},$$

while, by the proof of Theorem 3.1, the bias term is of order  $\tau$ . This result highlights the nontrivial interplay between the multiple length-scales  $\{\varepsilon_n^{(k)}\}_{k=1}^q$  and the regularization parameter  $\tau$ , which jointly determine the convergence behavior of the discrete minimizers. Understanding this relationship provides practical guidance for choosing these parameters to achieve an appropriate balance between bias, variance, and computational cost. Finally, the theorem generalizes known rates for graph-based learning: when  $q = 1$ , we recover the convergence rates established in [35, Corollary 1.8], thereby situating our result within and extending the existing theoretical framework on graphs.

*Remark 3.2* (Optimal regularization parameter  $\tau$ ). For fixed bandwidth parameters  $\varepsilon_n^{(k)}$ , the last two terms in (5) have the form  $A/\tau + B\tau$ , where  $A = (\varepsilon_n^{(1)})^{2p_1}$  and  $B = 1 + \sum_{k=1}^q \lambda_k \varepsilon_n^{(k)}$ . Minimizing this expression over  $\tau > 0$  yields the optimal choice  $\tau^* = \frac{(\varepsilon_n^{(1)})^{p_1}}{\sqrt{B}}$  and, substituting the latter in (5) gives

$$\|u_{n,\tau^*}^{(\mathbf{y}_n)} - g|_{\Omega_n}\|_{L^2(\mu_n)} \leq C \left[ \sum_{k=1}^q \lambda_k \frac{(\varepsilon_n^{(1)})^{2p_1}}{(\varepsilon_n^{(k)})^{2p_k}} \left( \frac{\log(n)}{n(\varepsilon_n^{(k)})^d} \right)^{1/2} + 2(\varepsilon_n^{(1)})^{p_1} \sqrt{1 + \sum_{k=1}^q \lambda_k \varepsilon_n^{(k)}} \right].$$

#### 3.4.2 Truncated energies

Next, we show that we can use the truncated version of HOHL in practice. In fact, going beyond heuristics, the below results shows that truncated energies converges to the same continuum energy as the full energy (see [87, Theorem 3.5]) This signifies that for large enough  $n$ , truncated and full energies will lead to arbitrarily close minimizers.

**Theorem 3.3** (Consistency of the truncated sum of Laplacians). *Assume that **S.1**, **M.1**, **M.2**, **W.1** and **D.1** hold. Let  $q \geq 1$ ,  $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$  with  $p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$ . Assume that  $\rho \in C^\infty$  and that  $\varepsilon_n^{(q)}$  satisfies*

$$\lim_{n \rightarrow \infty} \frac{\log(n)}{n \left(\varepsilon_n^{(q)}\right)^{d+4p_k}} = 0.$$

*Let  $R_n \leq n$  be a sequence with  $R_n \rightarrow \infty$ ,  $\Psi : \text{TL}^2(\Omega) \rightarrow \mathbb{R}$  a continuous function and  $(\mu_n, u_n)$  the minimizer of  $(\mathcal{S}\mathcal{J})_{n, E_n, \Psi, R_n}^{(q, P)}$ . Then,  $\mathbb{P}$ -a.e., there exists a subsequence  $(\mu_{n_k}, u_{n_k})$  converging to  $(\mu, u)$  in  $\text{TL}^2(\Omega)$  where  $(\mu, u)$  is a minimizer of  $(\mathcal{S}\mathcal{J})_{\infty, \Psi}^{(q, P)}$ .*

We emphasize that the convergence conditions we impose on the truncation are mild: it suffices that the truncation threshold tends to infinity. This grants practitioners considerable flexibility in applying HOHL in practice.

Moreover, we note from Lemma 4.6 that the limiting continuum energy in Theorem 3.3

$$\sum_{i=1}^{\infty} \left( \sum_{k=1}^q \lambda_k \beta_i^{p_k} \right) \langle v, \psi_i \rangle_{L^2(\mu)}^2$$

is equal to  $\langle v, \mathcal{L}^{(q)} v \rangle_{L^2(\mu)}$ . In particular, this implies that  $\mathcal{L}^{(q)}$  could also be defined directly through its spectrum: it shares the same eigenfunctions as  $\Delta_\rho$ , while its eigenvalues are given by functions of those of  $\Delta_\rho$ . This places our approach firmly within the framework of spectral kernel learning, where it is common to regularize with operators derived from the Laplacian. Spectral learning—and its analysis through reproducing kernel Hilbert space techniques—has been shown to yield powerful results for uncertainty quantification, enabling explicit bounds on expected error as well as estimates of prediction variance in semi-supervised learning (see [93] and references therein).

### 3.4.3 Non-geometrical setting

All of the preceding results focused on applying HOHL within the geometric setting. The following result extends the analysis to arbitrary hypergraphs, demonstrating that the matrix  $\mathcal{L}_{\text{dis}}^{(q)}$  can be interpreted as the Laplacian of a specially constructed graph (which may be signed [78]).

**Proposition 3.4.** *There exists a graph  $\tilde{G}$  whose Laplacian matrix is given by  $\mathcal{L}_{\text{dis}}^{(q)}$ . Furthermore,  $\mathcal{L}_{\text{dis}}^{(q)}$  is positive semi-definite and symmetric, and (2) is a quadratic form.*

This result is particularly noteworthy as it implies that standard numerical techniques developed for Laplace learning are directly applicable to HOHL. These include spectral truncation (see Theorem 3.3), Nyström extensions [29], conjugate gradient methods for Laplacian inversion, and more. Moreover, it suggests that HOHL can function as a drop-in replacement for Laplace learning within existing machine learning pipelines. To demonstrate this in practice, Section 5 presents active learning experiments where the HOHL matrix  $\mathcal{L}_n^{(q)}$  defines a Gaussian prior over functions.

We can extend the HOHL energy (2) to non-geometric datasets, where geometric embeddings for the vertices are unavailable and, for example, the weight models described in Section 3.2 do not apply. In such settings, the standard feature-based hypergraph construction, e.g. [46, 95], forms a hyperedge among all nodes that share a common categorical feature value. Each hyperedge is also assigned unit weight.

Unlike previous methods that rely on global hyperedge smoothing or iterative optimization, our approach introduces scalable, structure-aware regularization tailored to categorical feature data. Crucially, in contrast to the geometric setting, hyperedge size here does not reflect sample proximity but rather the frequency of shared attribute values. Large hyperedges correspond to common features and tend to encode coarse relationships, while small hyperedges capture more specific, and potentially more informative, structure. Promoting regularity over these smaller subsets is thus useful for fine-grained label propagation. This represents the inverse perspective of the geometric setting, where larger hyperedges encode finer local interactions. We summarize the main differences of HOHL in the geometric and non-geometric setting in Table 1.

Aspect	Geometric Setting	Non-Geometric Setting
Vertex set $V$	$\Omega_n = \{x_i\}_{i=1}^n \subset \mathbb{R}^d$	Arbitrary object set (no embedding in $\mathbb{R}^d$ )
Hyperedge construction	Based on distance/proximity (e.g., $\varepsilon$ -neighborhoods)	Based on shared attributes or features
Interpretation of hyperedge size	Smaller hyperedges correspond to longer-range geometric connections; larger hyperedges capture denser local neighborhoods	Smaller hyperedges reflect more specific or rare attributes; larger hyperedges correspond to common, broad features
Use of length scales $\varepsilon$	Essential for defining Laplacians $\Delta_{n,\varepsilon}$	Not applicable
HOHL regularization	Higher regularization on large hyperedges	Higher regularization on small hyperedges
Continuum limit of HOHL	$W^{p,q,2}$ semi-norm [87]	No natural continuum limit
Characterization of well/ill-posedness of HOHL in SSL	$\checkmark$ ([87, Theorem 3.5])	—
Rates of convergence for HOHL regularizer	$\checkmark$ (Theorem 3.1)	—
Use of spectral truncation	$\checkmark$ (consistency in Theorem 3.3)	$\checkmark$
HOHL is quadratic form	$\checkmark$	$\checkmark$

Table 1: Comparison of HOHL in geometric and non-geometric settings.

In real datasets however, even small hyperedges can contain many nodes, and large ones are common. This poses computational challenges for HOHL, which penalizes through powers of Laplacians on skeleton graphs. To address this, Algorithm 1 groups hyperedges by size and aggregates their skeleton graphs into a fixed number of levels. This reduces computational cost and imposes a multiscale hierarchy that prioritizes structurally meaningful interactions. In Section 5, we demonstrate that HOHL outperforms many other hypergraph methods in semi-supervised learning.

## 4 Proofs

In this section, we present the proofs of our results.

### 4.1 Fully supervised problem with HOHL regularization

For this section only, we proceed to a constant re-scaling of the Laplacians in Section 3.2. In particular, we define:

$$\Delta_{n,\varepsilon} = \frac{2}{n\varepsilon^{d+2}}(D_{n,\varepsilon} - W_{n,\varepsilon}) \quad \text{and} \quad \Delta_\rho u(x) = -\frac{\sigma_\eta}{\rho(x)} \operatorname{div}(\rho^2 \nabla u)(x).$$

We also recall that  $E^c$  denotes the complement of the set  $E$ .

First, the aim is to show the analogue of [35, Proposition 2.1] and to this purpose, we define

$$w_n = \left( \operatorname{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_{n,\varepsilon_n}^{p_k} \right)^{-1} \xi_n$$

---

**Algorithm 1** Construction of multiscale Laplacians for HOHL. Hyperedges are grouped by size, skeletons are aggregated into  $q$  segments, and Laplacians  $\{L^{(k)}\}$  are computed for use in (2).

---

**Input:** Hypergraph  $G = (V, E)$ ; number of skeleton graphs  $q$

**Output:** List of Laplacian matrices  $\{L^{(k)}\}_{k=1}^q$  to be used in (2)

---

```

1: Group hyperedges by size:  $A[j] \leftarrow \{e \in E : |e| = j\}$ 
2: Let  $\text{Ord} \leftarrow$  sorted list of unique hyperedge sizes (descending)
3: Initialize adjacency matrix list:  $\text{Adj} \leftarrow []$ 
4: for each  $j \in \text{Ord}$  do
5:   Construct skeleton graph from  $A[j]$  and append its adjacency matrix to  $\text{Adj}$ 
6: end for
7: Define uniform thresholds to split  $\text{Adj}$  into  $q$  segments and store them in the list  $\text{Thresholds}$ 
8: for each  $k = 1$  to  $q$  do
9:   Let  $\text{start}_k \leftarrow \text{Thresholds}[k - 1]$  ▷ First index of segment  $k$ 
10:  Let  $\text{end}_k \leftarrow \text{Thresholds}[k]$  ▷ One past the last index of segment  $k$ 
11:  Set  $W_n^{(k)} \leftarrow 0$ 
12:  for each  $m = \text{start}_k$  to  $\text{end}_k - 1$  do
13:     $W_n^{(k)} \leftarrow W_n^{(k)} + \text{Adj}[m]$ 
14:  end for
15:  Compute Laplacian  $L^{(k)}$  from  $W_n^{(k)}$ 
16: end for
17: return  $\{L^{(k)}\}_{k=1}^q$ 

```

---

as well as

$$(6) \quad \tilde{w}_n = \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \left( \frac{2}{n \binom{\varepsilon_n^{(k)}}{2}} D_{n, \varepsilon_n^{(k)}} \right)^{p_k} \right)^{-1} \xi_n$$

where  $\xi_n = (\xi_1, \dots, \xi_n)$  and  $D_{n, \varepsilon_n^{(k)}}$  is the diagonal degree matrix defined in Section 3.2.

**Lemma 4.1** (Bound on matrix product). *Assume that **S.1**, **M.1**, **M.2**, **D.1** and **W.2** hold. Furthermore, let  $\rho \in C^\infty$  and assume that  $W_{ii} = 0$ . Let  $\ell \in \mathbb{N}$ ,  $q \geq 1$ ,  $1 \leq k \leq q$ ,  $\{\lambda_r\}_{r=1}^q$  be a sequence of positive numbers,  $P = \{p_r\}_{r=1}^q \subseteq \mathbb{N}$  with  $1 \leq p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(r)}\}_{r=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)} > 0$ . Let  $\xi_i$  be iid, mean zero, sub-Gaussian random variables and  $\tilde{w}_n$  be defined in (6). Then, for  $\alpha > 1$ ,  $\tau > 0$  and  $\varepsilon_n^{(q)}$  satisfying*

$$\varepsilon_n^{(q)} \geq C \left( \frac{\log(n)}{n} \right)^{1/d},$$

*there exists  $C > 0$  such that*

$$(7) \quad \left\| W_{n, \varepsilon_n^{(k)}} D_{n, \varepsilon_n^{(k)}}^{\ell-1} \tilde{w}_n \right\|_{L^2(\mu_n)} \leq \frac{C n^\ell \left( \varepsilon_n^{(1)} \right)^{2p_1}}{\tau} \left( \frac{\log(n)}{n \binom{\varepsilon_n^{(k)}}{d}} \right)^{1/2}$$

*with probability  $1 - Cn^{-\alpha}$ .*

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

For notational convenience, we define  $d_{n, i, \varepsilon_n^{(r)}} = \sum_{j=1}^n \left( W_{n, \varepsilon_n^{(r)}} \right)_{ij}$ . For  $1 \leq r \leq q$ , we let  $E_r$  be the event where the graph  $G_n$  satisfies the following inequalities

- there exists constants  $C_1$  and  $C_2$  such that

$$(8) \quad C_1 \leq n^{-1} d_{n, i, \varepsilon_n^{(r)}} \leq C_2$$

for all  $1 \leq i \leq n$ ;

$$\bullet \#\{j \mid \left(W_{n,\varepsilon_n^{(r)}}\right)_{ij} > 0\} \leq Cn \left(\varepsilon_n^{(r)}\right)^d \text{ for } 1 \leq i \leq n.$$

Let  $E = \cap_{r=1}^q E_r$  be the set of events such that the above inequalities hold for all  $1 \leq r \leq q$ . By [35, Lemma 2.2], we know that  $\mathbb{P}(E_r) \geq 1 - 2ne^{-c(r)n\left(\varepsilon_n^{(r)}\right)^d}$ . Hence,

$$\mathbb{P}(E^c) \leq \sum_{r=1}^q \mathbb{P}(E_r^c) \leq \sum_{r=1}^q 2ne^{-c(r)n\left(\varepsilon_n^{(r)}\right)^d} \leq Cne^{-cn\left(\varepsilon_n^{(q)}\right)^d}$$

implying that

$$\mathbb{P}(E) \geq 1 - Cne^{-cn\left(\varepsilon_n^{(q)}\right)^d}.$$

Now, let  $G_n$  be a graph in the event  $E$  and fix  $1 \leq i \leq n$ . For  $1 \leq j \leq n$ , let

$$q_j^i = \frac{\tau \left(W_{n,\varepsilon_n^{(k)}}\right)_{ij} \left(d_{n,j,\varepsilon_n^{(k)}}\right)^{\ell-1} \xi_j}{1 + \tau \sum_{r=1}^q \lambda_r \left(\frac{2}{n\left(\varepsilon_n^{(r)}\right)^2} d_{n,j,\varepsilon_n^{(r)}}\right)^{p_r}}$$

and we note that

$$(9) \quad \tau \left(W_{n,\varepsilon_n^{(k)}} D_{n,\varepsilon_n^{(k)}}^{\ell-1} \tilde{w}_n\right)_i = \sum_{j=1}^n q_j^i.$$

Now,  $q_j^i$  are centered and independent random variables. Furthermore, we estimate as follows:

$$(10) \quad \frac{1}{n^{\ell-1} \left(\varepsilon_n^{(1)}\right)^{2p_1}} |q_j^i| = \tau |\xi_j| \left(W_{n,\varepsilon_n^{(k)}}\right)_{ij} \frac{\left(d_{n,j,\varepsilon_n^{(k)}}\right)^{\ell-1}}{n^{\ell-1}} \frac{1}{\left(\varepsilon_n^{(1)}\right)^{2p_1}} \frac{1}{1 + \tau \sum_{r=1}^q \lambda_r \left(\frac{2}{n\left(\varepsilon_n^{(r)}\right)^2} d_{n,j,\varepsilon_n^{(r)}}\right)^{p_r}}$$

$$(11) \quad \leq \frac{C|\xi_j|}{\left(\varepsilon_n^{(k)}\right)^d} \frac{1}{\left(\varepsilon_n^{(1)}\right)^{2p_1}} \frac{1}{\sum_{r=1}^q \lambda_r \frac{1}{\left(\varepsilon_n^{(r)}\right)^{2p_r}}}$$

where we used the fact that  $\left(W_{n,\varepsilon_n^{(k)}}\right)_{ij} \leq C \left(\varepsilon_n^{(k)}\right)^{-d}$  and (8) for (10) and the fact that  $p_1 \leq \dots \leq p_q$  and  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$  for (11). This implies that  $[n^{\ell-1}(\varepsilon_n^{(1)})]^{-1} q_j^i$  are sub-Gaussian and satisfy the same inequalities in the Birnbaum-Orlicz norm as in [35, Lemma 2.4]. By applying the same Hoeffding inequality as in the latter, for any  $t > 0$ , we therefore obtain

$$\mathbb{P} \left( \frac{1}{n^{\ell-1} \left(\varepsilon_n^{(1)}\right)^{2p_1}} \left| \sum_{j=1}^n q_j^i \right| > t \mid E \right) \leq 2e^{-ct^2 \left(\varepsilon_n^{(k)}\right)^d / n}.$$

We then choose  $t = \lambda \sqrt{\frac{n \log(n)}{\left(\varepsilon_n^{(k)}\right)^d}}$  so that, using (9),

$$(12) \quad \frac{\tau}{n^{\ell-1} \left(\varepsilon_n^{(1)}\right)^{2p_1}} \left| \left(W_{n,\varepsilon_n^{(k)}} D_{n,\varepsilon_n^{(k)}}^{\ell-1} \tilde{w}_n\right)_i \right| = \frac{\tau}{n^{\ell-1} \left(\varepsilon_n^{(1)}\right)^{2p_1}} \left| \sum_{j=1}^n q_j^i \right| \leq \lambda \sqrt{\frac{n \log(n)}{\left(\varepsilon_n^{(k)}\right)^d}}$$

with probability at least  $1 - 2n^{-c\lambda^2}$  conditioned on  $E$ . We pick  $\lambda = \sqrt{\frac{\alpha+1}{c}}$  and, through an union bound, obtain that (12) holds for all  $1 \leq i \leq n$  with probability at least  $1 - 2n^{1-c\lambda^2} = 1 - 2n^{-\alpha}$ , conditioned on  $E$ . Starting from (12), we get

$$(13) \quad \left\| W_{n, \varepsilon_n^{(k)}} D_{n, \varepsilon_n^{(k)}}^{\ell-1} \tilde{w}_n \right\|_{L^2(\mu_n)} \leq \left\| W_{n, \varepsilon_n^{(k)}} D_{n, \varepsilon_n^{(k)}}^{\ell-1} \tilde{w}_n \right\|_{L^\infty(\mu_n)} \leq \frac{C n^\ell \left( \varepsilon_n^{(1)} \right)^{2p_1}}{\tau} \left( \frac{\log(n)}{n \left( \varepsilon_n^{(k)} \right)^d} \right)^{1/2}$$

conditioned on  $E$  with probability at least  $1 - Cn^{-\alpha}$ . Let  $A$  be the event such that (7) holds. By (13),

$$\mathbb{P}(A) = \mathbb{P}(A | E) \mathbb{P}(E) + \mathbb{P}(A | E^c) \mathbb{P}(E^c) \geq (1 - Cn^{-\alpha}) \cdot \left( 1 - Cn e^{-cn \left( \varepsilon_n^{(q)} \right)^d} \right)$$

and, to conclude, we can pick  $C$  large enough so that  $\mathbb{P}(A) \geq 1 - Cn^{-\alpha}$ .  $\square$

**Lemma 4.2** (Bounds on  $\tilde{w}_n$ ). *Assume that S.1, M.1, M.2, D.1 and W.2 hold. Furthermore, let  $\rho \in C^\infty$  and assume that  $W_{ii} = 0$ . Let  $q \geq 1$ ,  $\{\lambda_k\}_{k=1}^q$  be a sequence of positive numbers,  $P = \{p_k\}_{k=1}^q \subseteq \mathbb{N}$  with  $1 \leq p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)} > 0$ . Let  $\xi_i$  be iid, mean zero, sub-Gaussian random variables and  $\tilde{w}_n$  be defined in (6). Then, for all  $\alpha > 1$ , there exists  $\varepsilon_0 > 0$  and  $C > 0$  such that for all  $E_n$  satisfying*

$$\varepsilon_0 \geq \varepsilon_n^{(1)} \geq \dots \geq \varepsilon_n^{(q)} \geq C \left( \frac{\log(n)}{n} \right)^{1/d},$$

and  $\tau > 0$ , the following holds with probability  $1 - Cn^{-\alpha}$ :

1.

$$(14) \quad \left\| \frac{1}{2} \nabla \mathcal{R}_{n, \tau}^{(\xi_n)}(\tilde{w}_n) \right\|_{L^2(\mu_n)} \leq C \sum_{k=1}^q \lambda_k \frac{\left( \varepsilon_n^{(1)} \right)^{2p_1}}{\left( \varepsilon_n^{(k)} \right)^{2p_k}} \left( \frac{\log(n)}{n \left( \varepsilon_n^{(k)} \right)^d} \right)^{1/2};$$

2.

$$(15) \quad \|\tilde{w}_n\|_{L^2(\mu_n)} \leq \frac{C}{\tau} \left( \varepsilon_n^{(1)} \right)^{2p_1}.$$

*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. Let  $\|\cdot\|_{\text{op}}$  denote the operator norm.

We start by noting that

$$(16) \quad \frac{1}{2} \nabla \mathcal{R}_{n, \tau}^{(\mathbf{a}_n)}(v_n) = v_n - \mathbf{a}_n + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} v_n = \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} \right) v_n - \mathbf{a}_n.$$

In particular, this implies that, with probability at least  $1 - Cn^{-\alpha}$  (see below), we can estimate as follows:

$$(17) \quad \left\| \frac{1}{2} \nabla \mathcal{R}_{n, \tau}^{(\xi_n)}(\tilde{w}_n) \right\|_{L^2(\mu_n)} = \left\| \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} \right) \tilde{w}_n - \xi_n \right\|_{L^2(\mu_n)}$$

$$(18) \quad = \left\| \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} \right) \tilde{w}_n - \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \left( \frac{2}{n \left( \varepsilon_n^{(k)} \right)^2} D_{n, \varepsilon_n^{(k)}} \right)^{p_k} \right) \tilde{w}_n \right\|_{L^2(\mu_n)} \\ \leq C \tau \sum_{k=1}^q \frac{\lambda_k}{n^{p_k} \left( \varepsilon_n^{(k)} \right)^{2p_k}} \left\| \left[ \left( D_{n, \varepsilon_n^{(k)}} - W_{n, \varepsilon_n^{(k)}} \right)^{p_k} - D_{n, \varepsilon_n^{(k)}}^{p_k} \right] \tilde{w}_n \right\|_{L^2(\mu_n)}$$

$$(19) \quad = C\tau \sum_{k=1}^q \frac{\lambda_k}{n^{p_k} \left(\varepsilon_n^{(k)}\right)^{2p_k}} \left\| \left[ \left( \sum_{\chi \in \{0,1\}^{p_k}} \prod_{i=1}^{p_k} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i} \right) - D_{n,\varepsilon_n^{(k)}}^{p_k} \right] \tilde{w}_n \right\|_{L^2(\mu_n)}$$

where we used (16) in (17), (6) in (18), and the expansion

$$(D_{n,\varepsilon_n^{(k)}} - W_{n,\varepsilon_n^{(k)}})^{p_k} = \sum_{\chi \in \{0,1\}^{p_k}} \prod_{i=1}^{p_k} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i}$$

for (19). Subtracting the term  $D_{n,\varepsilon_n^{(k)}}^{p_k}$  from  $\left( \sum_{\chi \in \{0,1\}^{p_k}} \prod_{i=1}^{p_k} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i} \right)$  removes the summand associated with  $\chi = (1, 1, \dots, 1)$ , so that every remaining product in the sum contains at least one factor of  $W_{n,\varepsilon_n^{(k)}}$  and

$$\left( \sum_{\chi \in \{0,1\}^{p_k}} \prod_{i=1}^{p_k} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i} \right) - D_{n,\varepsilon_n^{(k)}}^{p_k} = \sum_{\substack{\chi \in \{0,1\}^{p_k} \\ \chi \neq (1,\dots,1)}} \prod_{i=1}^{p_k} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i}$$

For any fixed  $\chi \in \{0,1\}^{p_k}$  with  $\chi \neq (1, 1, \dots, 1)$ , let  $r_\chi$  denote the index of the first occurrence of  $W_{n,\varepsilon_n^{(k)}}$  when reading the product from right to left (the index exists since all terms with  $\chi \neq (1, 1, \dots, 1)$  contain at least one factor of  $W_{n,\varepsilon_n^{(k)}}$ ). We can then factor the product as

$$\prod_{i=1}^{p_k} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i} = \underbrace{\left( \prod_{i=1}^{p_k-r_\chi} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i} \right)}_{=: T_{-r_\chi}} (-W_{n,\varepsilon_n^{(k)}})^{r_\chi-1} D_{n,\varepsilon_n^{(k)}}^{r_\chi-1}.$$

The term  $T_{-r_\chi}$  contains  $p_k - r_\chi$  factors, each equal to either  $D_{n,\varepsilon_n^{(k)}}$  or  $W_{n,\varepsilon_n^{(k)}}$ . Using the operator-norm bounds from [35, Lemma 2.3], we have  $\|T_{-r_\chi}\|_{\text{op}} \leq (Cn)^{p_k-r_\chi}$ . This implies that

$$\begin{aligned} & \left\| \left[ \left( \sum_{\chi \in \{0,1\}^{p_k}} \prod_{i=1}^{p_k} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i} \right) - D_{n,\varepsilon_n^{(k)}}^{p_k} \right] \tilde{w}_n \right\|_{L^2(\mu_n)} \\ &= \left\| \left[ \sum_{\substack{\chi \in \{0,1\}^{p_k} \\ \chi \neq (1,\dots,1)}} \prod_{i=1}^{p_k} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i} \right] \tilde{w}_n \right\|_{L^2(\mu_n)} \\ &\leq \sum_{\substack{\chi \in \{0,1\}^{p_k} \\ \chi \neq (1,\dots,1)}} \left\| \prod_{i=1}^{p_k} D_{n,\varepsilon_n^{(k)}}^{\chi_i} (-W_{n,\varepsilon_n^{(k)}})^{1-\chi_i} \tilde{w}_n \right\|_{L^2(\mu_n)} \\ &\leq \sum_{\substack{\chi \in \{0,1\}^{p_k} \\ \chi \neq (1,\dots,1)}} \|T_{-r_\chi}\|_{\text{op}} \left\| W_{n,\varepsilon_n^{(k)}}^{r_\chi-1} D_{n,\varepsilon_n^{(k)}}^{r_\chi-1} \tilde{w}_n \right\|_{L^2(\mu_n)} \\ &\leq C \sum_{\substack{\chi \in \{0,1\}^{p_k} \\ \chi \neq (1,\dots,1)}} n^{p_k-r_\chi} \frac{n^{r_\chi} \left(\varepsilon_n^{(1)}\right)^{2p_1}}{\tau} \left( \frac{\log(n)}{n \left(\varepsilon_n^{(k)}\right)^d} \right)^{1/2} \end{aligned} \tag{20}$$

$$(21) \quad = C(2^{p_k} - 1) \frac{n^{p_k} \left(\varepsilon_n^{(1)}\right)^{2p_1}}{\tau} \left( \frac{\log(n)}{n \left(\varepsilon_n^{(k)}\right)^d} \right)^{1/2}$$



where we used Lemma 4.1 for (20). Inserting (21) into (19), we obtain

$$\begin{aligned}
\left\| \frac{1}{2} \nabla \mathcal{R}_{n,\tau}^{(\xi_n)}(\tilde{w}_n) \right\|_{L^2(\mu_n)} &\leq C \tau \sum_{k=1}^q \frac{\lambda_k}{n^{p_k} \left( \varepsilon_n^{(k)} \right)^{2p_k}} \frac{n^{p_k} \left( \varepsilon_n^{(1)} \right)^{2p_1}}{\tau} \left( \frac{\log(n)}{n \left( \varepsilon_n^{(k)} \right)^d} \right)^{1/2} \\
(22) \qquad &= C \sum_{k=1}^q \frac{\lambda_k \left( \varepsilon_n^{(1)} \right)^{2p_1}}{\left( \varepsilon_n^{(k)} \right)^{2p_k}} \left( \frac{\log(n)}{n \left( \varepsilon_n^{(k)} \right)^d} \right)^{1/2}.
\end{aligned}$$

For the second claim of the lemma, let us start by assuming that  $G_n$  is a graph in the event  $E$  from the proof of Lemma 4.1. Then,

$$\begin{aligned}
\|\tilde{w}_n\|_{L^2(\mu_n)}^2 &= \frac{1}{n} \sum_{i=1}^n \frac{\xi_i^2}{\left( 1 + \tau \sum_{r=1}^q \lambda_r \left( \frac{2}{n \left( \varepsilon_n^{(r)} \right)^2} d_{n,i,\varepsilon_n^{(r)}} \right)^{p_r} \right)^2} \\
(23) \qquad &\leq \frac{C}{n} \sum_{i=1}^n \frac{\xi_i^2}{\left( \tau \sum_{r=1}^q \frac{\lambda_r}{\left( \varepsilon_n^{(r)} \right)^{2p_r}} \right)^2} \\
(24) \qquad &\leq \frac{C}{n\tau^2} \left( \varepsilon_n^{(1)} \right)^{4p_1} \sum_{i=1}^n \xi_i^2
\end{aligned}$$

where we used the fact that there exists  $C_1 \leq n^{-1} d_{n,i,\varepsilon_n^{(r)}} \leq C_2$  for all  $1 \leq i \leq n$  and  $1 \leq r \leq q$  for (23) and the fact that  $p_1 \leq \dots \leq p_q$  and  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$  for (24). Let  $A$  be the event such that (15) holds. Then, arguing as in [36, Lemma 2.6], we can show that  $\mathbb{P}(A | E) \geq 1 - Cn^{-\alpha}$ . Analogously to the proof of Lemma 4.1, we conclude that  $\mathbb{P}(A) \geq 1 - Cn^{-\alpha}$ .  $\square$

**Proposition 4.3** (Rates between discrete noisy and noiseless minimizers). *Assume that **S.1**, **M.1**, **M.2**, **D.1** and **W.2** hold. Furthermore, let  $\rho \in C^\infty$  and assume that  $W_{ii} = 0$ . Let  $q \geq 1$ ,  $\{\lambda_k\}_{k=1}^q$  be a sequence of positive numbers,  $P = \{p_k\}_{k=1}^q \subseteq \mathbb{N}$  with  $1 \leq p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)} > 0$ . Let  $\xi_i$  be iid, mean zero, sub-Gaussian random variables,  $g \in C^\infty$ ,  $\mathbf{y}_n = \{y_i\}_{i=1}^n$  with  $y_i = g(x_i) + \xi_i$  and  $\mathbf{g}_n = \{g(x_i)\}_{i=1}^n$ . Then, for all  $\alpha > 1$ , there exists  $\varepsilon_0 > 0$  and  $C > 0$  such that for all  $E_n$  satisfying*

$$\varepsilon_0 \geq \varepsilon_n^{(1)} \geq \dots \geq \varepsilon_n^{(q)} \geq C \left( \frac{\log(n)}{n} \right)^{1/d},$$

and  $\tau > 0$ , the following holds with probability  $1 - Cn^{-\alpha}$ :

$$\|u_{n,\tau}^{(\mathbf{y}_n)} - u_{n,\tau}^{(\mathbf{g}_n)}\|_{L^2(\mu_n)} \leq C \left( \sum_{k=1}^q \lambda_k \frac{\left( \varepsilon_n^{(1)} \right)^{2p_1}}{\left( \varepsilon_n^{(k)} \right)^{2p_k}} \left( \frac{\log(n)}{n \left( \varepsilon_n^{(k)} \right)^d} \right)^{1/2} + \frac{\left( \varepsilon_n^{(1)} \right)^{2p_1}}{\tau} \right)$$

where  $u_{n,\tau}^{(\mathbf{y}_n)}$  and  $u_{n,\tau}^{(\mathbf{g}_n)}$  are the minimizers of  $\mathcal{R}_{n,\tau}^{(\mathbf{y}_n)}$  and  $\mathcal{R}_{n,\tau}^{(\mathbf{g}_n)}$  respectively.

*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  $v_n^{(1)}, v_n^{(2)} : \Omega \mapsto \mathbb{R}$ , we start by estimating as follows using (16):

$$\begin{aligned}
\left\langle \frac{1}{2} \nabla \mathcal{R}_{n,\tau}^{(\mathbf{a}_n)}(v_n^{(1)}) - \frac{1}{2} \nabla \mathcal{R}_{n,\tau}^{(\mathbf{a}_n)}(v_n^{(2)}), v_n^{(1)} - v_n^{(2)} \right\rangle_{L^2(\mu_n)} &= \|v_n^{(1)} - v_n^{(2)}\|_{L^2(\mu_n)}^2 \\
&+ \tau \sum_{k=1}^q \lambda_k \left\langle \Delta_{n,\varepsilon_n^{(k)}}^{p_k} \left( v_n^{(1)} - v_n^{(2)} \right), v_n^{(1)} - v_n^{(2)} \right\rangle_{L^2(\mu_n)}.
\end{aligned}$$

Since  $\Delta_{n, \varepsilon_n}^{p_k}$  is positive semi-definite, using the Cauchy-Schwarz inequality, we can conclude that

$$\|v_n^{(1)} - v_n^{(2)}\|_{L^2(\mu_n)} \leq \frac{1}{2} \|\nabla \mathcal{R}_{n, \tau}^{(\mathbf{a}_n)}(v_n^{(1)}) - \nabla \mathcal{R}_{n, \tau}^{(\mathbf{a}_n)}(v_n^{(2)})\|_{L^2(\mu_n)}.$$

Furthermore, by first order optimality and (16), we have that

$$u_{n, \tau}^{(\mathbf{y}_n)} - \mathbf{y}_n + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n}^{p_k} u_{n, \tau}^{(\mathbf{y}_n)} = 0$$

and

$$u_{n, \tau}^{(\mathbf{g}_n)} - \mathbf{g}_n + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n}^{p_k} u_{n, \tau}^{(\mathbf{g}_n)} = 0$$

implying that

$$(25) \quad u_{n, \tau}^{(\mathbf{y}_n)} - u_{n, \tau}^{(\mathbf{g}_n)} + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n}^{p_k} (u_{n, \tau}^{(\mathbf{y}_n)} - u_{n, \tau}^{(\mathbf{g}_n)}) = \xi_n$$

or equivalently

$$(26) \quad u_{n, \tau}^{(\mathbf{y}_n)} - u_{n, \tau}^{(\mathbf{g}_n)} = \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n}^{p_k} \right)^{-1} \xi_n = w_n.$$

We can now estimate as follows, with probability  $1 - Cn^{-\alpha}$  (see below):

$$(27) \quad \|u_{n, \tau}^{(\mathbf{y}_n)} - u_{n, \tau}^{(\mathbf{g}_n)}\|_{L^2(\mu_n)} \leq \|w_n - \tilde{w}_n\|_{L^2(\mu_n)} + \|\tilde{w}_n\|_{L^2(\mu_n)} \\ \leq \frac{1}{2} \|\nabla \mathcal{R}_{n, \tau}^{(\xi_n)}(w_n) - \nabla \mathcal{R}_{n, \tau}^{(\xi_n)}(\tilde{w}_n)\|_{L^2(\mu_n)} + \|\tilde{w}_n\|_{L^2(\mu_n)}$$

$$(28) \quad \leq C \left( \|\nabla \mathcal{R}_{n, \tau}^{(\xi_n)}(\tilde{w}_n)\|_{L^2(\mu_n)} + \frac{(\varepsilon_n^{(1)})^{2p_1}}{\tau} \right)$$

$$(29) \quad \leq C \left( \sum_{k=1}^q \lambda_k \frac{(\varepsilon_n^{(1)})^{2p_1}}{(\varepsilon_n^{(k)})^{2p_k}} \left( \frac{\log(n)}{n (\varepsilon_n^{(k)})^d} \right)^{1/2} + \frac{(\varepsilon_n^{(1)})^{2p_1}}{\tau} \right)$$

where we used (26) for (27), the fact that  $\nabla \mathcal{R}_{n, \tau}^{(\xi_n)}(w_n) = 0$  and (15) for (28) as well as (14) for (29).  $\square$

**Proposition 4.4** (Rates between discrete noiseless and continuum minimizers). *Assume that **S.1**, **M.1**, **M.2**, **D.1** and **W.2** hold. Furthermore, let  $\rho \in C^\infty$  and assume that  $W_{ii} = 0$ . Let  $q \geq 1$ ,  $\{\lambda_k\}_{k=1}^q$  be a sequence of positive numbers,  $P = \{p_k\}_{k=1}^q \subseteq \mathbb{N}$  with  $1 \leq p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)} > 0$ . Let  $\xi_i$  be iid, mean zero, sub-Gaussian random variables,  $g \in C^\infty$  and  $\mathbf{g}_n = \{g(x_i)\}_{i=1}^n$ . Then, for all  $\alpha > 1$  and  $\tau_0$ , there exists  $\varepsilon_0 > 0$  and  $C > c > 0$  such that for all  $E_n$  satisfying*

$$\varepsilon_0 \geq \varepsilon_n^{(1)} \geq \dots \geq \varepsilon_n^{(q)} \geq C \left( \frac{\log(n)}{n} \right)^{1/d},$$

and  $0 < \tau < \tau_0$ , the following holds with probability  $1 - Cn^{-\alpha} - Cne^{-cn(\varepsilon_n^{(q)})^{d+4p_q}}$ :

$$(30) \quad \|u_\tau|_{\Omega_n} - u_{n, \tau}^{(\mathbf{g}_n)}\|_{L^2(\mu_n)} \leq C\tau \sum_{k=1}^q \lambda_k \varepsilon_n^{(k)}.$$

where  $u_{n, \tau}^{(\mathbf{g}_n)}$  and  $u_\tau$  are the minimizers of  $\mathcal{R}_{n, \tau}^{(\mathbf{g}_n)}$  and  $\mathcal{R}_{\infty, \tau}^{(g)}$  respectively.

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

We start the proof by proving the following fact: if  $w_n$  satisfies

$$(31) \quad \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} \right) v_n = \mathbf{a}_n,$$

then  $\|v_n\|_{L^2(\mu_n)} \leq \|\mathbf{a}_n\|_{L^2(\mu_n)}$ . Indeed, by Proposition 3.4, we know that  $\sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k}$  is a graph Laplacian, so we can apply the same proof as in [35, Lemma 2.14] with the eigenpairs of  $\mathcal{L}_n^{(q)} = \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k}$  to deduce (31).

Next, by first order conditions, we note that  $u_\tau$  satisfies the equivalent continuum identity

$$(32) \quad \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_\rho^{p_k} \right) u_\tau - g = 0$$

from which we deduce that

$$(33) \quad \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} \right) u_\tau - g = \tau \left( \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} - \sum_{k=1}^q \lambda_k \Delta_\rho^{p_k} \right) u_\tau.$$

We then estimate as follows:

$$(34) \quad \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} \right) (u_\tau|_{\Omega_n} - u_{n, \tau}^{(\mathbf{g}_n)}) = \left( \text{Id} + \tau \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} \right) u_\tau|_{\Omega_n} - \mathbf{g}_n$$

$$(35) \quad = \tau \left( \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} - \sum_{k=1}^q \lambda_k \Delta_\rho^{p_k} \right) u_\tau|_{\Omega_n}$$

where we used the fact that  $u_{n, \tau}^{(\mathbf{g}_n)}$  satisfies (31) with  $\mathbf{a}_n = \mathbf{g}_n$  by first order conditions for (34) and where we used (33) (as well as a slight abuse of notation) for (35).

Let  $E_k$  be the event such that [35, Theorem 2.8] holds for  $\varepsilon_n^{(k)}$ : we have

$$\mathbb{P}(E_k) \geq 1 - Cn^{-\alpha} - Cne^{-cn(\varepsilon_n^{(k)})^{d+4p_k}}$$

which implies that

$$\mathbb{P} \left( \bigcup_{k=1}^q E_k^c \right) \leq \sum_{k=1}^q \mathbb{P}(E_k^c) \leq Cn^{-\alpha} + Cne^{-cn(\varepsilon_n^{(q)})^{d+4p_q}}$$

where we used the fact that  $p_1 \leq \dots \leq p_q$  and  $\{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$  for the last inequality. In turn, this means that

$$\mathbb{P} \left( \bigcap_{k=1}^q E_k \right) \geq 1 - Cn^{-\alpha} - Cne^{-cn(\varepsilon_n^{(q)})^{d+4p_q}}.$$

We therefore obtain, with probability at least  $1 - Cn^{-\alpha} - Cne^{-cn(\varepsilon_n^{(q)})^{d+4p_q}}$ :

$$(36) \quad \|u_\tau|_{\Omega_n} - u_{n, \tau}^{(\mathbf{g}_n)}\|_{L^2(\mu_n)} \leq \left\| \tau \left( \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}^{p_k} - \sum_{k=1}^q \lambda_k \Delta_\rho^{p_k} \right) u_\tau \right\|_{L^2(\mu)}$$

$$(37) \quad \leq \tau \sum_{k=1}^q \lambda_k \left\| \left( \Delta_{n, \varepsilon_n^{(k)}}^{p_k} - \Delta_\rho^{p_k} \right) u_\tau \right\|_{L^2(\mu)} \\ \leq C\tau \sum_{k=1}^q \lambda_k \varepsilon_n^{(k)} (1 + \|u_\tau\|_{C^{2p_k+1}})$$

where we used the fact that  $u_\tau|_{\Omega_n} - u_{n,\tau}^{(\mathbf{g}_n)}$  satisfies (31) with  $\mathbf{a}_n = \tau \left( \sum_{k=1}^q \lambda_k \Delta_{n,\varepsilon_n}^{p_k} - \sum_{k=1}^q \lambda_k \Delta_\rho^{p_k} \right) u_\tau|_{\Omega_n}$  for (36) and [35, Theorem 2.8] for (37).

To establish the desired result, it remains to verify that  $\sup_{0 < \tau < \tau_0} \|u_\tau\|_{C^{2p_k+1}} \leq C$ . To that end, we start by noting that (32) implies that

$$\begin{aligned} \langle g, \psi_i \rangle_{L^2(\mu)} &= \langle u_\tau, \psi_i \rangle_{L^2(\mu)} + \tau \sum_{k=1}^q \lambda_k \langle \Delta_\rho^{p_k} u_\tau, \psi_i \rangle_{L^2(\mu)} \\ (38) \quad &= \langle u_\tau, \psi_i \rangle_{L^2(\mu)} + \tau \sum_{k=1}^q \lambda_k \beta_i^{p_k} \langle u_\tau, \psi_i \rangle_{L^2(\mu)} \end{aligned}$$

$$(39) \quad = \langle u_\tau, \psi_i \rangle_{L^2(\mu)} \left( 1 + \tau \sum_{k=1}^q \lambda_k \beta_i^{p_k} \right)$$

where we used the fact that  $\Delta_\rho$  is self-adjoint for (38). Then, for  $s > 0$ , we compute as follows:

$$\begin{aligned} \|u_\tau\|_{\mathcal{H}^s(\Omega)}^2 &= \sum_{i=1}^{\infty} \beta_i^s \langle u_\tau, \psi_i \rangle_{L^2(\mu)}^2 \\ (40) \quad &= \sum_{i=1}^{\infty} \beta_i^s \frac{\langle g, \psi_i \rangle_{L^2(\mu)}^2}{\left( 1 + \tau \sum_{k=1}^q \lambda_k \beta_i^{p_k} \right)^2} \\ &\leq \sum_{i=1}^{\infty} \beta_i^s \langle g, \psi_i \rangle_{L^2(\mu)}^2 \\ &= \|g\|_{\mathcal{H}^s(\Omega)}^2 \end{aligned}$$

where we used (39) for (40). By [23, Lemma 17], there exists  $c$  and  $C$  such that  $c\|h\|_{W^{s,2}(\Omega)} \leq \|h\|_{\mathcal{H}^s(\Omega)} \leq C\|h\|_{W^{s,2}(\Omega)}$  for all  $h \in \mathcal{H}^s(\Omega)$ . From the above, we therefore deduce that  $\|u_\tau\|_{W^{s,2}(\Omega)} \leq C\|g\|_{W^{s,2}(\Omega)}$ . Finally, by Morrey's inequality [55], for  $s$  sufficiently large there exists  $C' > 0$  such that

$$\|u_\tau\|_{C^{2p_k+1}} \leq C' \|u_\tau\|_{W^{s,2}} \leq C' \|g\|_{W^{s,2}(\Omega)}.$$

Since  $g \in C^\infty(\Omega)$ , taking the supremum of  $\tau$  over  $(0, \tau_0)$  concludes the proof.  $\square$

*Proof of Theorem 3.1.* 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 with an estimate between the continuum solution  $u_\tau$  and  $g$ . Similarly to (16), it can easily be verified that

$$\frac{1}{2} \nabla \mathcal{R}_\infty^{(g)}(v) = v - g + \tau \sum_{k=1}^q \lambda_k \Delta_\rho^{p_k} v$$

from which we deduce the following identity

$$(41) \quad \langle \nabla \mathcal{R}_\infty^{(g)}(w), w - v \rangle_{L^2(\mu)} - \|w - v\|_{L^2(\mu)}^2 - \tau \sum_{k=1}^q \lambda_k \langle w - v, \Delta_\rho^{p_k}(w - v) \rangle_{L^2(\mu)} = \mathcal{R}_\infty^{(g)}(w) - \mathcal{R}_\infty^{(g)}(v).$$

for any  $w, v \in W^{p_q,2}$ . Then, we have

$$(42) \quad \|u_\tau - g\|_{L^2(\mu)}^2 + \tau \sum_{k=1}^q \lambda_k \langle u_\tau - g, \Delta_\rho^{p_k}(u_\tau - g) \rangle_{L^2(\mu)} = \mathcal{R}_\infty^{(g)}(g) - \mathcal{R}_\infty^{(g)}(u_\tau)$$

$$(43) \quad = \langle \nabla \mathcal{R}_\infty^{(g)}(g), g - u_\tau \rangle_{L^2(\mu)} - \|g - u_\tau\|_{L^2(\mu)}^2 - \tau \sum_{k=1}^q \lambda_k \langle g - u_\tau, \Delta_\rho^{p_k}(g - u_\tau) \rangle_{L^2(\mu)}$$

where we used (41) for (42) with  $w = u_\tau$ ,  $v = g$  and (41) for (43) with  $w = g$ ,  $v = u_\tau$ . We can therefore conclude that

$$\|u_\tau - g\|_{L^2(\mu)}^2 \leq \frac{1}{2} \|\nabla \mathcal{R}_\infty^{(g)}(g)\|_{L^2(\mu)} \|g - u_\tau\|_{L^2(\mu)}$$

or equivalently

$$(44) \quad \|u_\tau - g\|_{L^2(\mu)} \leq \tau \sum_{k=1}^q \lambda_k \langle g, \Delta_\rho^{p_k} g \rangle_{L^2(\mu)} \leq C\tau.$$

We now combine all the previous rates:

$$\begin{aligned} \|u_{n,\tau}^{(\mathbf{y}_n)} - g|_{\Omega_n}\|_{L^2(\mu_n)} &= \|u_{n,\tau}^{(\mathbf{y}_n)} - u_{n,\tau}^{(\mathbf{g}_n)}\|_{L^2(\mu_n)} + \|u_{n,\tau}^{(\mathbf{g}_n)} - u_\tau|_{\Omega_n}\|_{L^2(\mu_n)} + \|u_\tau|_{\Omega_n} - g|_{\Omega_n}\|_{L^2(\mu_n)} \\ &=: T_1 + T_2 + T_3. \end{aligned}$$

We can bound  $T_1$  using Proposition 4.3 and  $T_2$  using Proposition 4.4. For  $T_3$ , we proceed as follows. Let  $T_n : \Omega_n \rightarrow \Omega$  be a transport map satisfying  $(T_n)_\# \mu = \mu_n$ . Then, we have

$$\begin{aligned} \|u_\tau|_{\Omega_n} - g|_{\Omega_n}\|_{L^2(\mu_n)} &= \|u_\tau|_{\Omega_n} \circ T_n - g|_{\Omega_n} \circ T_n\|_{L^2(\mu)} \\ &\leq \|u_\tau|_{\Omega_n} \circ T_n - u_\tau\|_{L^2(\mu)} + \|u_\tau - g\|_{L^2(\mu)} + \|g - g|_{\Omega_n} \circ T_n\|_{L^2(\mu)} \\ &=: T_4 + T_5 + T_6. \end{aligned}$$

Since  $g \in C^\infty(\Omega)$ ,  $g$  is Lipschitz and

$$(45) \quad T_6 \leq C\|T_n - \text{Id}\|_{L^2(\mu)}.$$

Similarly, from the proof of Proposition 4.4, we recall that  $\sup_{0 < \tau < \tau_0} \|u_\tau\|_{C^{2p_k+1}(\Omega)} \leq C$  which implies that  $u_\tau$  is bounded in  $C^1(\Omega)$  and hence Lipschitz. Consequently, we can bound

$$(46) \quad T_4 \leq C\|T_n - \text{Id}\|_{L^2(\mu)}.$$

Since the choice of  $T_n$  is arbitrary among all maps satisfying  $(T_n)_\# \mu = \mu_n$ , we take the optimal one minimizing  $\|T_n - \text{Id}\|_{L^2(\mu)}$ . By the probabilistic transport bound of [28], this distance satisfies

$$\|T_n - \text{Id}\|_{L^2(\mu)} \leq C \left( \frac{|\log(\delta)|}{n} \right)^{1/d}$$

with probability at least  $1 - \delta$ . By picking  $\delta = n^{-\alpha}$ , combining (45), (46) and (44) for  $T_5$ , we obtain

$$T_3 \leq C \left( \tau + \left( \frac{\log(n)}{n} \right)^{1/d} \right)$$

with probability  $1 - n^{-\alpha}$  which concludes the proof.  $\square$

## 4.2 Truncated energies

For clarity of presentation, we divide the proof of Theorem 3.3 into two parts. We begin by establishing the result in the simpler case  $q = 1$ . Next, we examine the spectral convergence properties of the operator  $\mathcal{L}_n^{(q)}$ , and by incorporating this analysis into the  $q = 1$  argument, we obtain the general case.

### 4.2.1 Convergence of truncated energies in the single Laplacian case

The aim of this section is to prove the following result which corresponds to Theorem 3.3 when  $q = 1$ . For notational simplicity, we make the following assumption on the length scale  $\varepsilon_n$ .

*Assumptions 5.* Assumptions on the length-scale.

**L.1** The length scale  $\varepsilon = \varepsilon_n$  is positive, converges to 0, i.e.  $0 < \varepsilon_n \rightarrow 0$  and satisfies the following lower bound:

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

**Proposition 4.5.** Assume that **S.1**, **M.1**, **M.2**, **D.1** and **W.2** hold. Let  $s > 0$  and  $\varepsilon_n$  satisfy **L.1**. Let  $K_n \leq n$  be a sequence with  $K_n \rightarrow \infty$ ,  $\Psi : \text{TL}^2(\Omega) \rightarrow \mathbb{R}$  a continuous function and  $(\mu_n, u_n)$  the minimizer of  $(\mathcal{S}\mathcal{J})_{n, \{\varepsilon_n\}, \Psi, R_n}^{(1, \{s\})}$ . Then,  $\mathbb{P}$ -a.e., there exists a subsequence  $(\mu_{n_k}, u_{n_k})$  converging to  $(\mu, u)$  in  $\text{TL}^2(\Omega)$  where  $(\mu, u)$  is a minimizer of  $(\mathcal{S}\mathcal{J})_{\infty, \Psi}^{(1, \{s\})}$ .

*Proof.* 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.

Our aim is to show that the functionals  $(\mathcal{S}\mathcal{J})_{n, \{\varepsilon_n\}, \Psi, K_n}^{(1, \{s\})}$   $\Gamma$ -converge to  $(\mathcal{S}\mathcal{J})_{\infty, \Psi}^{(1, \{s\})}$  and satisfy the compactness property. Once we can do this, all conditions from Proposition 2.6 are satisfied and we can conclude.

Let us define the functionals

$$(\mathcal{S}\mathcal{J})_{n, \{\varepsilon_n\}, K_n}^{(1, \{s\})}((\nu, v)) = \begin{cases} \sum_{k=1}^{K_n} a_{n, \varepsilon_n, k}^s \langle \phi_{n, \varepsilon_n, k}, v \rangle_{L^2(\mu_n)}^2 & \text{if } \nu = \mu_n \text{ and } \langle \psi_{n, k}, v \rangle_n = 0 \text{ for all } k > K_n \\ \infty & \text{else} \end{cases}$$

where  $a_{n, \varepsilon_n, k}$  are the eigenvalues of  $\Delta_{n, \varepsilon_n}$ , and

$$(\mathcal{S}\mathcal{J})_{\infty}((\nu, v)) = \begin{cases} \sum_{k=1}^{\infty} \beta_k^s \langle \psi_k, v \rangle_{L^2(\mu)}^2 & \text{if } \nu = \mu \\ \infty & \text{else.} \end{cases}$$

The latter functionals are similar to  $(\mathcal{S}\mathcal{J})_{n, \{\varepsilon_n\}, \Psi, K_n}^{(1, \{s\})}$  and  $(\mathcal{S}\mathcal{J})_{\infty, \Psi}^{(1, \{s\})}$ , the only difference being that they do not contain the data fidelity term  $\Psi$ .

First, we tackle the  $\liminf$ -inequality. We assume that  $(\nu, v) \in \text{TL}^2(\Omega)$  and that  $(\nu_n, v_n) \rightarrow (\nu, v)$  in  $\text{TL}^2$ . If  $\liminf_{n \rightarrow \infty} (\mathcal{S}\mathcal{J})_{n, \{\varepsilon_n\}, K_n}^{(1, \{s\})}((\nu_n, v_n)) = +\infty$ , then the inequality is trivial. Hence, without loss of generality, let us assume that  $\sup_{n \in \mathbb{N}} (\mathcal{S}\mathcal{J})_{n, \{\varepsilon_n\}, K_n}^{(1, \{s\})}((\nu_n, v_n)) \leq C$ . In particular, this implies that  $\nu_n = \mu_n$ ,  $\langle \phi_{n, \varepsilon_n, k}, v_n \rangle_{L^2(\mu_n)} = 0$  for all  $k > K_n$  and  $\sup_{n \in \mathbb{N}} \sum_{k=1}^{K_n} a_{n, \varepsilon_n, k}^s \langle \phi_{n, \varepsilon_n, k}, v \rangle_{L^2(\mu_n)}^2 \leq C$ . Since we have  $\mu_n \rightarrow \nu$  weakly (by the  $\text{TL}^2$ -convergence assumption—see Proposition 2.2) and  $\mu_n \rightarrow \mu$  weakly (convergence of the empirical measures), we conclude (by the uniqueness of weak limits) that  $\nu = \mu$ . We then proceed as in [23, Theorem 2.2].

Let us start by assuming that  $\sum_{k=1}^{\infty} \beta_k^s \langle v, \psi_k \rangle_{L^2(\Omega)}^2 < \infty$ . In particular, since  $\phi_{n, \varepsilon_n, k} \rightarrow \psi_k$  and  $v_n \rightarrow v$  in  $\text{TL}^2(\Omega)$  [23], we have that  $\langle \phi_{n, \varepsilon_n, k}, v_n \rangle_{L^2(\mu_n)} \rightarrow \langle v, \psi_k \rangle_{L^2(\mu)}$  by [37, Proposition 2.6]. Furthermore, by [37, Theorem 1.2], we have  $a_{n, \varepsilon_n, k} \rightarrow \beta_k$ . Now, let  $\delta > 0$  and pick  $K$  such that

$$\sum_{k=1}^K \beta_k^s \langle v, \psi_k \rangle_{L^2(\Omega)}^2 \geq \sum_{k=1}^{\infty} \beta_k^s \langle v, \psi_k \rangle_{L^2(\Omega)}^2 - \delta.$$

Since  $K_n \rightarrow \infty$ , we have

$$\begin{aligned} \liminf_{n \rightarrow \infty} \sum_{k=1}^{K_n} a_{n, \varepsilon_n, k}^s \langle \phi_{n, \varepsilon_n, k}, v \rangle_{L^2(\mu_n)}^2 &\geq \liminf_{n \rightarrow \infty} \sum_{k=1}^K a_{n, \varepsilon_n, k}^s \langle \phi_{n, \varepsilon_n, k}, v \rangle_{L^2(\mu_n)}^2 \\ &= \sum_{k=1}^K \beta_k^s \langle v, \psi_k \rangle_{L^2(\Omega)}^2 \\ &\geq \sum_{k=1}^{\infty} \beta_k^s \langle v, \psi_k \rangle_{L^2(\Omega)}^2 - \delta. \end{aligned}$$

Taking  $\delta \rightarrow 0$ , we obtain the  $\liminf$ -inequality. Now, assume that  $\sum_{k=1}^{\infty} \beta_k^s \langle v, \psi_k \rangle_{L^2(\Omega)}^2 = \infty$ . Then, for any  $K \in \mathbb{N}$ , we have

$$\begin{aligned} C &\geq \liminf_{n \rightarrow \infty} \sum_{k=1}^{K_n} a_{n, \varepsilon_n, k}^s \langle \phi_{n, \varepsilon_n, k}, v \rangle_{L^2(\mu_n)}^2 \\ &\geq \lim_{K \rightarrow \infty} \liminf_{n \rightarrow \infty} \sum_{k=1}^K a_{n, \varepsilon_n, k}^s \langle \phi_{n, \varepsilon_n, k}, v \rangle_{L^2(\mu_n)}^2 \end{aligned}$$

$$\begin{aligned}
&= \lim_{K \rightarrow \infty} \sum_{k=1}^K \beta_k^s \langle v, \psi_k \rangle_{L^2(\Omega)}^2 \\
&= \infty
\end{aligned}$$

which is a contradiction.

For the lim sup-inequality, we let  $(\nu, v) \in \text{TL}^2(\Omega)$ . If  $(\mathcal{ST})_{n, \{\varepsilon_n\}, K_n}^{(1, \{s\})}((\nu, v)) = \infty$ , the inequality is trivial, so we assume that  $\nu = \mu$  and  $\sum_{k=1}^{\infty} \beta_k^s \langle \psi_k, v \rangle_{L^2(\mu)}^2 < \infty$  or equivalently  $v \in W^{s,2}(\Omega)$  [23]. If we can prove the lim sup-inequality on a dense subset of  $\{\mu\} \times W^{s,2}(\Omega)$ , namely  $\{\mu\} \times C_c^\infty(\Omega)$ , we can conclude due to [36, Remark 2.7].

Let  $v \in C_c^\infty(\Omega)$  and define  $v_n$  to be the restriction of  $v$  to  $\Omega_n$ . Let us consider the sequence  $(\mu_n, \bar{v}_n)$  where  $\bar{v}_n = v_n - \sum_{k=K_n+1}^n \langle v_n, \phi_{n,\varepsilon_n,k} \rangle_n \phi_{n,\varepsilon_n,k}$ . It is clear that  $\langle \phi_{n,\varepsilon_n,k}, \bar{v}_n \rangle_n = 0$  for  $k > K_n$ . We now verify that  $(\mu_n, \bar{v}_n) \rightarrow (\mu, v)$  in  $\text{TL}^2(\Omega)$ . With  $T_n$  the transport maps of Theorem 2.3, we estimate as follows:

$$\begin{aligned}
\int_{\Omega} |\bar{v}_n \circ T_n - v|^2 d\mu &\leq 2 \underbrace{\int_{\Omega} |v_n \circ T_n - v|^2 d\mu}_{=: T_1} + 2 \int_{\Omega} |\bar{v}_n \circ T_n - v_n \circ T_n|^2 d\mu \\
&\leq 2T_1 + 2 \sum_{i=1}^n \langle \bar{v}_n - v_n, \phi_{n,\varepsilon_n,k} \rangle_{L^2(\mu_n)}^2 \\
(47) \quad &\leq 2T_1 + \frac{2}{a_{n,\varepsilon_n,K_n+1}^2} \sum_{k=K_n+1}^n a_{n,\varepsilon_n,k}^2 \langle v_n, \phi_{n,\varepsilon_n,k} \rangle_{L^2(\mu_n)}^2
\end{aligned}$$

$$(48) \quad \leq 2T_1 + \frac{C}{a_{n,\varepsilon_n,K_n+1}^2}$$

where we used the fact that the eigenvalues are ordered for (47) and [89, Lemma 4.19] for (48). We know from [37, Theorem 1.4] that  $T_1 \rightarrow 0$  and from the proof of [23, Theorem 2.2] that  $a_{n,\varepsilon_n,K_n+1}^2 \rightarrow \infty$  which allows us to conclude that  $(\mu_n, \bar{v}_n) \rightarrow (\mu, v)$  in  $\text{TL}^2(\Omega)$ .

Since  $v \in C_c^\infty$  then  $v \in W^{m,2}$  for any  $m \in \mathbb{N}$ . Choose  $m \in \mathbb{N}$  with  $m > \frac{s}{2}$  and let  $\delta > 0$  be such that  $s + \delta = 2m$ . As an intermediary step, let us compute:

$$\begin{aligned}
T_2 &:= \sum_{k=1}^n a_{n,\varepsilon_n,k}^s \left\langle \phi_{n,\varepsilon_n,k}, \sum_{j=K_n+1}^n \langle v_n, \phi_{n,\varepsilon_n,j} \rangle_{L^2(\mu_n)} \phi_{n,\varepsilon_n,j} \right\rangle_{L^2(\mu_n)}^2 \\
&= \sum_{k=K_n+1}^n a_{n,\varepsilon_n,k}^s \langle \phi_{n,\varepsilon_n,k}, v_n \rangle_{L^2(\mu_n)}^2 \\
&\leq \frac{1}{a_{n,\varepsilon_n,K_n+1}^\delta} \sum_{k=K_n+1}^n a_{n,\varepsilon_n,k}^{s+\delta} \langle \phi_{n,\varepsilon_n,k}, v_n \rangle_{L^2(\mu_n)}^2 \\
&\leq \frac{C}{a_{n,\varepsilon_n,K_n+1}^\delta}.
\end{aligned}$$

Arguing as above, we obtain that  $T_2 \rightarrow 0$ . We conclude by estimating as follows:

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \sqrt{\sum_{k=1}^{K_n} a_{n,\varepsilon_n,k}^s \langle \phi_{n,\varepsilon_n,k}, \bar{v}_n \rangle_{L^2(\mu_n)}^2} &\leq \limsup_{n \rightarrow \infty} \sqrt{\sum_{k=1}^n a_{n,\varepsilon_n,k}^s \langle \phi_{n,\varepsilon_n,k}, \bar{v}_n \rangle_{L^2(\mu_n)}^2} \\
(49) \quad &\leq \limsup_{n \rightarrow \infty} \sqrt{\sum_{k=1}^n a_{n,\varepsilon_n,k}^s \langle \phi_{n,\varepsilon_n,k}, v_n \rangle_{L^2(\mu_n)}^2} + \limsup_{n \rightarrow \infty} \sqrt{T_2}
\end{aligned}$$

$$(50) \quad \leq \sqrt{\sum_{k=1}^{\infty} \beta_k^s \langle \psi_k, v \rangle_{L^2(\mu)}^2}$$

where used [89, Lemma 4.15] for (49), [89, Proposition 4.21] and the fact that  $T_2 \rightarrow 0$  for (50). Squaring the last inequality, we obtain the lim sup-inequality.

Summarizing the above two results, we obtain that  $(\mathcal{S}\mathcal{J})_{n,\{\varepsilon_n\},\Psi,K_n}^{(1,\{s\})}$   $\Gamma$ -converges to  $(\mathcal{S}\mathcal{J})_{\infty,\Psi}^{(1,\{s\})}$ . Since  $\Psi$  is continuous in  $\text{TL}^2(\Omega)$ , we use Proposition 2.7 to deduce that

$$(\mathcal{S}\mathcal{J})_{n,\{\varepsilon_n\},\Psi,K_n}^{(1,\{s\})} \quad \Gamma\text{-converges to} \quad (\mathcal{S}\mathcal{J})_{\infty,\Psi}^{(1,\{s\})}.$$

Let us now consider a sequence  $(\mu_n, v_n)$  minimizing  $(\mathcal{S}\mathcal{J})_{n,\{\varepsilon_n\},\Psi,K_n}^{(1,\{s\})}$  with  $\sup_{n \in \mathbb{N}} \|v_n\|_{L^2(\mu_n)} \leq C$ . In particular, we note that  $\langle \phi_{n,\varepsilon_n,k}, v_n \rangle_n = 0$  for all  $k > K_n$  and recall that  $K_n \rightarrow \infty$ . Therefore, we can apply the same proof as in [23, Theorem 2.2] to show that there exists a converging subsequence in  $\text{TL}^2(\Omega)$ .

Specifically,  $\sup_{n \in \mathbb{N}} \|v_n\|_{L^2(\mu_n)} \leq C$  implies that  $\sup_{n \in \mathbb{N}} \sum_{k=1}^{K_n} \langle v_n, \phi_{n,\varepsilon_n,k} \rangle_{L^2(\mu_n)}^2 \leq C$ . Hence, by a diagonal procedure, we can find a sequence  $n_m \rightarrow \infty$  such that for every  $k$ ,  $\langle v_{n_m}, \psi_{n_m,\varepsilon_{n_m},k} \rangle_{L^2(\mu_{n_m})}$  converges to some coefficient  $\gamma_k$ . By Fatou's lemma,  $\sum_{k=1}^{\infty} |\gamma_k|^2 \leq \liminf_{m \rightarrow \infty} \sum_{k=1}^{n_m} |\langle v_{n_m}, \phi_{n_m,\varepsilon_{n_m},k} \rangle_{L^2(\mu_{n_m})}|^2 \leq C$ , so we can define  $v = \sum_{k=1}^{\infty} \gamma_k \psi_k \in L^2(\mu)$ . Using [23, Lemma 7.7], we obtain a sequence  $R_{n_m} \rightarrow \infty$  such that  $\sum_{k=1}^{R_{n_m}} \langle v_{n_m}, \phi_{n_m,\varepsilon_{n_m},k} \rangle_{L^2(\mu_{n_m})} \phi_{n_m,\varepsilon_{n_m},k} \rightarrow v$  in  $\text{TL}^2(\Omega)$ . We note that  $R_{n_m}$  can always be picked such that  $R_{n_m} \leq K_{n_m}$ . Indeed, the  $\text{TL}^2$ -convergence resulting from [23, Lemma 7.7] holds for any sequence converging to  $\infty$  and majorized by  $R_{n_m}$ : therefore, we can always pick  $R_{n_m} = \min\{R_{n_m}, K_{n_m}\}$ .

Then, we check the convergence in  $\text{TL}^2$  of  $v_{n_m}$  to  $v$ :

$$\begin{aligned} \|v_{n_m} \circ T_{n_m} - v\|_{L^2(\mu)} &\leq \|v_{n_m} - \sum_{k=1}^{R_{n_m}} \langle v_{n_m}, \phi_{n_m,\varepsilon_{n_m},k} \rangle_{L^2(\mu_{n_m})} \phi_{n_m,\varepsilon_{n_m},k}\|_{L^2(\mu_n)} \\ &\quad + \left\| \sum_{k=1}^{R_{n_m}} \langle v_{n_m}, \phi_{n_m,\varepsilon_{n_m},k} \rangle_{L^2(\mu_{n_m})} \phi_{n_m,\varepsilon_{n_m},k} \circ T_{n_m} - v \right\|_{L^2(\mu)} \\ &\leq \frac{1}{a_{n_m,\varepsilon_{n_m},R_{n_m}}^s} \sum_{k=R_{n_m}+1}^{K_{n_m}} a_{n_m,\varepsilon_{n_m},k}^s \langle v_{n_m}, \phi_{n_m,\varepsilon_{n_m},k} \rangle_{L^2(\mu_{n_m})}^2 \phi_{n_m,\varepsilon_{n_m},k} \\ &\quad + \left\| \sum_{k=1}^{R_{n_m}} \langle v_{n_m}, \phi_{n_m,\varepsilon_{n_m},k} \rangle_{L^2(\mu_{n_m})} \phi_{n_m,\varepsilon_{n_m},k} \circ T_{n_m} - v \right\|_{L^2(\mu)} \\ &\leq \frac{C}{a_{n_m,\varepsilon_{n_m},R_{n_m}}^s} + \left\| \sum_{k=1}^{R_{n_m}} \langle v_{n_m}, \phi_{n_m,\varepsilon_{n_m},k} \rangle_{L^2(\mu_{n_m})} \phi_{n_m,\varepsilon_{n_m},k} \circ T_{n_m} - v \right\|_{L^2(\mu)} \end{aligned}$$

where the last inequality follows from the fact that  $\sup_{n \in \mathbb{N}} \sum_{k=1}^{K_n} a_{n,\varepsilon_n,k}^s \langle v_n, \phi_{n,\varepsilon_n,k} \rangle_{L^2(\mu_n)}^2 \leq C$  since  $(\mu_n, v_n)$  are minimizers of  $(\mathcal{S}\mathcal{J})_{n,\{\varepsilon_n\},\Psi,K_n}$  (see also [89, Lemma 4.25]). In order to conclude that  $v_{n_m} \rightarrow v$  in  $\text{TL}^2(\Omega)$ , we note the following two facts: the first term in the last inequality tends to 0 as argued in [23, Theorem 2]; the second term tends to 0 since  $\sum_{k=1}^{R_{n_m}} \langle v_{n_m}, \psi_{n_m,k} \rangle_{L^2(\mu_{n_m})} \psi_{n_m,k} \rightarrow v$  in  $\text{TL}^2(\Omega)$ . By Proposition 2.6, we know that the limiting point  $v$  is a minimizer of  $(\mathcal{S}\mathcal{J})_{\infty,\Psi}^{(1,\{s\})}$ .  $\square$

#### 4.2.2 Spectral convergence of $\mathcal{L}_n^{(q)}$

In this section, we analyze the spectral convergence of  $\mathcal{L}_n^{(q)}$  and, by combining with the results of the previous section, prove Theorem 3.3.

**Lemma 4.6** (Eigenpairs of  $\mathcal{L}^{(q)}$ ). *Assume that S.1, M.1, M.2, W.1 and D.1 hold. Let  $q \geq 1$ ,  $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$  with  $p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$ . Assume that  $\rho \in C^\infty$ . The eigenpairs of  $\mathcal{L}^{(q)}$  are  $\{(\sum_{k=1}^q \lambda_k \beta_i^{p_k}, \psi_i)\}_{i=1}^\infty$ .*

*Proof.* First, we see that

$$\mathcal{L}^{(q)} \psi_i = \sum_{k=1}^q \lambda_k \Delta_\rho^{p_k} \psi_i = \sum_{k=1}^q \lambda_k \beta_i^{p_k} \psi_i.$$

This implies that  $(\sum_{k=1}^q \lambda_k \beta_i^{p_k}, \psi_i)$  is an eigenpair of  $\mathcal{L}^{(q)}$ . We now consider two cases.



**Case 1.** Assume  $(\sum_{k=1}^q \lambda_k \beta_j^{p_k}, \psi)$  is an eigenpair of  $\mathcal{L}^{(q)}$ . Then,

$$\begin{aligned} \mathcal{L}^{(q)} \psi &= \mathcal{L}^{(q)} \sum_{i=1}^{\infty} \langle \psi_i, \psi \rangle_{L^2(\mu)} \psi_i \\ &= \sum_{i=1}^{\infty} \langle \psi_i, \psi \rangle_{L^2(\mu)} \left( \sum_{k=1}^q \lambda_k \beta_i^{p_k} \right) \psi_i \end{aligned}$$

and

$$\mathcal{L}^{(q)} \psi = \sum_{i=1}^{\infty} \langle \psi_i, \psi \rangle_{L^2(\mu)} \left( \sum_{k=1}^q \lambda_k \beta_j^{p_k} \right) \psi_i.$$

Hence,

$$\sum_{i=1}^{\infty} \langle \psi_i, \psi \rangle_{L^2(\mu)} \psi_i \left( \sum_{k=1}^q \lambda_k \beta_i^{p_k} - \sum_{k=1}^q \lambda_k \beta_j^{p_k} \right) = 0.$$

Since,  $\{\psi_i\}_{i=1}^{\infty}$  are linearly independent then  $\langle \psi_i, \psi \rangle_{L^2(\mu)} \left( \sum_{k=1}^q \lambda_k \beta_i^{p_k} - \sum_{k=1}^q \lambda_k \beta_j^{p_k} \right) = 0$  for all  $i \in \mathbb{N}$ . Hence for  $i \neq j$  (since  $\beta_i \neq \beta_j$ ) we have  $\langle \psi_i, \psi \rangle_{L^2(\mu)} = 0$ . As  $\|\psi\|_{L^2(\mu)} = 1$  (assuming we normalised) then  $\psi = \pm \psi_j$ .

**Case 2.** Assume  $(\beta, \psi)$  is an eigenpair of  $\mathcal{L}^{(q)}$ . Then an analogous calculation to the one above implies

$$\sum_{i=1}^{\infty} \langle \psi_i, \psi \rangle_{L^2(\mu)} \left( \beta - \sum_{k=1}^q \lambda_k \beta_i^{p_k} \right) \psi_i = 0.$$

Again, as  $\{\psi_i\}_{i=1}^{\infty}$  are linearly independent then  $\langle \psi_i, \psi \rangle_{L^2(\mu)} \left( \beta - \sum_{k=1}^q \lambda_k \beta_i^{p_k} \right) = 0$  for all  $i \in \mathbb{N}$ . Since  $\psi \neq 0$ , then at least one  $\langle \psi_i, \psi \rangle_{L^2(\mu)} \neq 0$ . For this  $i$  we then must have  $\beta = \sum_{k=1}^q \lambda_k \beta_i^{p_k}$  and so we are back in Case 1.  $\square$

**Proposition 4.7** (Convergence of eigenpairs). *Assume that **S.1**, **M.1**, **M.2**, **W.1** and **D.1** hold. Let  $q \geq 1$ ,  $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$  with  $p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$ . Assume that  $\rho \in C^\infty$ . Assume that  $\varepsilon_n^{(q)}$  satisfies **L.1**. Then,  $\mathbb{P}$ -a.s., the following holds:*

1.  $\beta_{n,i} \rightarrow \sum_{k=1}^q \lambda_k \beta_i^{p_k}$ ;
2.  $(\mu_n, \psi_{n,i}) \rightarrow (\mu, \psi_i)$  in  $\text{TL}^2(\Omega)$ .

*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.

In order to prove the proposition, we want to proceed as in [37, Theorem 1.2] where the authors show the analogous result for a single Laplacian matrix  $\Delta_{n, \varepsilon_n}$ . In particular, the proof relies on the following results:

1.  $\Delta_{n, \varepsilon_n}$  and  $\Delta_\rho$  are self-adjoint and positive semi-definite;
2. the functional  $\langle v, \Delta_{n, \varepsilon_n} v \rangle_{L^2(\mu_n)}$   $\Gamma$ -converges to  $\langle v, \Delta_\rho v \rangle_{L^2(\mu)}$  [37, Theorem 1.4];
3. if a sequence satisfies  $\sup_n \langle v, \Delta_{n, \varepsilon_n} v \rangle_{L^2(\mu_n)} \leq C$  and  $\|v_n\|_{L^2(\mu_n)} \leq C$ , then there exists a converging subsequence in  $\text{TL}^2(\Omega)$  [37, Theorem 1.4].

Our Laplacian  $\mathcal{L}_n^{(q)}$  satisfies the same three properties:

1. Since each  $\Delta_{n, \varepsilon_n^{(k)}}$  is self-adjoint and positive semi-definite, so is  $\mathcal{L}_n^{(q)} = \sum_{k=1}^q \lambda_k \Delta_{n, \varepsilon_n^{(k)}}$ . The same argument applies to  $\mathcal{L}^{(q)}$ .
2. The fact that  $\langle v, \mathcal{L}_n^{(q)} v \rangle_{L^2(\mu_n)}$   $\Gamma$ -converges to  $\langle v, \mathcal{L}^{(q)} v \rangle_{L^2(\mu)}$  was shown in the proof of [87, Theorem 3.5].

3. If we assume that a sequence satisfies  $\sup_n \langle v_n, \mathcal{L}_n^{(q)} v_n \rangle_{L^2(\mu_n)} \leq C$  and  $\|v_n\|_{L^2(\mu_n)} \leq C$ , then in particular  $\sup_n \langle v_n, \Delta_{n, \varepsilon_n}^{p_1(q)} v_n \rangle_{L^2(\mu_n)} \leq C$  and  $\|v_n\|_{L^2(\mu_n)} \leq C$ : we can therefore use [23, Theorem 2] to deduce the existence of a converging subsequence in  $TL^2(\Omega)$ .

Specifically, let us start with the eigenvalues. First, we recall that since  $\mathcal{L}_n^{(q)}$  is self-adjoint and positive semi-definite, we can apply the Courant-Fisher characterization of eigenvalues [16, Max-min theorem] to infer that

$$(51) \quad \beta_{n,i} = \sup_{S \in \Sigma_{n,i-1}} \min_{v \in S^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)}$$

where  $\Sigma_{n,i-1}$  denotes the subspaces of  $\mathbb{R}^n$  of dimension  $i-1$  and  $S^\perp$  denotes the orthogonal complement of  $S$  with respect to the inner product in  $L^2(\mu_n)$ . We now proceed by induction on  $i$ .

**Base case  $i = 1$ .** We first observe that the graphs  $(\Omega_n, W_{n, \varepsilon_n^{(k)}})$  are connected  $\mathbb{P}$ -a.e. for  $n$  large enough. This follows from the ordering  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$  and from Assumption **(L.1)**, which guarantees connectivity in the random geometric graph regime [43, 66]. Consequently, for all sufficiently large  $n$ , the first eigenpair  $(0, \mathbf{1})$ , where  $\mathbf{1} \in \mathbb{R}^n$  is the constant-one vector, is shared across all Laplacians  $\Delta_{n, \varepsilon_n^{(k)}}$ . Thus, the first eigenpair of the discrete operator  $\mathcal{L}_n^{(q)}$  is given by  $(\beta_{n,1}, \psi_{n,1}) = (0, \mathbf{1})$ .

Furthermore, since the domain  $\Omega$  is connected by Assumption **S.1**, the continuum Laplacian  $\Delta_\rho$  has first eigenpair  $(0, \mathbf{1})$ , where  $\mathbf{1}$  denotes the constant function equal to one. By Lemma 4.6, the first eigenpair of the continuum limit operator  $\mathcal{L}^{(q)}$  is

$$\left( \sum_{k=1}^q \lambda_k \beta_1^{p_k}, \psi_1 \right) = (0, \mathbf{1}).$$

It follows that  $\beta_{n,1} \rightarrow \beta_1 = 0$  and  $(\mu_n, \psi_{n,1}) \rightarrow (\mu, \psi_1)$  in  $TL^2(\Omega)$  is satisfied.

**Induction step.** Now, suppose that  $\beta_{n,\ell} \rightarrow \beta_\ell$  for all  $\ell \leq i-1$ .

*Proof of the lower bound.* Let  $S \in \Sigma_{i-1}$ , where  $\Sigma_{i-1}$  denotes the subspaces of  $L^2(\Omega)$  of dimension  $i-1$ . In this case, we will also write  $S^\perp$  for the orthogonal complement of  $S$  with respect to the inner product in  $L^2(\mu)$ . Let  $\{v_1, \dots, v_{i-1}\}$  be an orthonormal basis of  $S$ . For each  $\ell = 1, \dots, i-1$ , the lim sup-inequality in [87, Theorem 3.5] ensures the existence of a sequence of functions  $v_{n,\ell} \in L^2(\mu_n)$  such that  $(\mu_n, v_{n,\ell}) \rightarrow (\mu, v_\ell)$  in  $TL^2(\Omega)$  as  $n \rightarrow \infty$ . By [37, Proposition 2.6], we have for all  $1 \leq \ell \leq i-1$ ,

$$\lim_{n \rightarrow \infty} \|v_{n,\ell}\|_{L^2(\mu_n)} = \|v_\ell\|_{L^2(\mu)} = 1,$$

and for all  $\ell \neq j$ ,

$$(52) \quad \lim_{n \rightarrow \infty} \langle v_{n,\ell}, v_{n,j} \rangle_{L^2(\mu_n)} = \langle v_\ell, v_j \rangle_{L^2(\mu)} = 0.$$

These results guarantee that for sufficiently large  $n$ , the set  $\{v_{n,1}, \dots, v_{n,i-1}\}$  spans a  $(i-1)$ -dimensional subspace of  $L^2(\mu_n)$ . We can then apply the Gram-Schmidt orthonormalization process to obtain an orthonormal basis  $\{\tilde{v}_{n,1}, \dots, \tilde{v}_{n,i-1}\}$ . Namely, we define

$$\tilde{v}_{n,1} := \frac{v_{n,1}}{\|v_{n,1}\|_{L^2(\mu_n)}},$$

and recursively for  $\ell = 2, \dots, i-1$ ,

$$\tilde{h}_{n,\ell} := v_{n,\ell} - \sum_{j=1}^{i-1} \langle v_{n,\ell}, \tilde{v}_{n,j} \rangle_{L^2(\mu_n)} \tilde{v}_{n,j}, \quad \tilde{v}_{n,\ell} := \frac{\tilde{h}_{n,\ell}}{\|\tilde{h}_{n,\ell}\|_{L^2(\mu_n)}}.$$

By (52) and [37, Proposition 2.6], it is straight-forward to check that  $\tilde{v}_{n,\ell} \rightarrow v_\ell$  in  $TL^2(\Omega)$  for  $1 \leq \ell \leq i-1$ . Let  $S_n \in \Sigma_{n,i-1}$  be the subset spanned by  $\{\tilde{v}_{n,1}, \dots, \tilde{v}_{n,i-1}\}$ .

We now want to show that

$$(53) \quad \liminf_{n \rightarrow \infty} \beta_{n,i} \geq \min_{v \in S, \|v\|_{L^2(\mu)}=1} \langle v, \mathcal{L}^{(q)} v \rangle_{L^2(\mu)}.$$

First, by (51), since  $\beta_{n,i} \geq \min_{v \in S_n^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)}$ , if

$$\liminf_{n \rightarrow \infty} \min_{v \in S_n^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)} = \infty,$$

then (53) is trivially satisfied. We therefore consider the case when

$$\liminf_{n \rightarrow \infty} \min_{v \in S_n^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)} < \infty$$

and, without loss of generality (see [37, 89]), we can assume that

$$\liminf_{n \rightarrow \infty} \min_{v \in S_n^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)} = \lim_{n \rightarrow \infty} \min_{v \in S_n^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)} < \infty.$$

Let  $w_n \in S_n^\perp$  be a sequence such that  $\|w_n\|_{L^2(\mu_n)} = 1$  and

$$\lim_{n \rightarrow \infty} \langle \mathcal{L}_n^{(q)} w_n, w_n \rangle_{L^2(\mu_n)} = \lim_{n \rightarrow \infty} \min_{v \in S_n^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)} < \infty.$$

Since  $\lim_{n \rightarrow \infty} \langle \mathcal{L}_n^{(q)} w_n, w_n \rangle_{L^2(\mu_n)} < \infty$ , we have that  $\sup_n \langle \mathcal{L}_n^{(q)} w_n, w_n \rangle_{L^2(\mu_n)} < \infty$  and, in particular,

$$\sup_n \langle \Delta_{n, \varepsilon_n^{(1)}}^{p_1} w_n, w_n \rangle_{L^2(\mu_n)} < \infty.$$

By [23, Theorem 2], we therefore obtain a converging subsequence  $(\mu_{n_m}, w_{n_m}) \rightarrow (\mu, w)$  in  $\text{TL}^2(\Omega)$ . By [37, Proposition 2.6], we deduce that  $\|w\|_{L^2(\mu)} = \lim_{m \rightarrow \infty} \|w_{n_m}\|_{L^2(\mu_{n_m})} = 1$ . Furthermore, since  $w_{n_m} \in S_{n_m}^\perp$  and  $\tilde{v}_{n_m, \ell} \rightarrow v_\ell$ , we also have  $\langle w, v_\ell \rangle_{L^2(\mu)} = \lim_{m \rightarrow \infty} \langle w_{n_m}, \tilde{v}_{n_m, \ell} \rangle_{L^2(\mu_{n_m})} = 0$  for  $1 \leq \ell \leq i-1$ , which implies that  $w \in S^\perp$ . Combining the latter facts about  $w$ , we estimate as follows:

$$(54) \quad \begin{aligned} \min_{v \in S^\perp, \|v\|_{L^2(\mu)}=1} \langle v, \mathcal{L}^{(q)} v \rangle_{L^2(\mu)} &\leq \langle w, \mathcal{L}^{(q)} w \rangle_{L^2(\mu)} \\ &\leq \liminf_{m \rightarrow \infty} \langle w_{n_m}, \mathcal{L}_{n_m}^{(q)} w_{n_m} \rangle_{L^2(\mu_{n_m})} \\ &= \lim_{n \rightarrow \infty} \min_{v \in S_n^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)} \\ &\leq \liminf_{n \rightarrow \infty} \sup_{\bar{S} \in \Sigma_{n, i-1}} \min_{v \in \bar{S}^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)} \\ (55) \quad &= \liminf_{n \rightarrow \infty} \beta_{n,i} \end{aligned}$$

where we used the  $\liminf$ -inequality of [87, Theorem 3.5] for (54) and (51) for (55). Finally, taking the supremum of all  $S \in \Sigma_{i-1}$  in (53), applying the Courant-Fisher characterization to the self-adjoint and positive semi-definite operator  $\mathcal{L}^{(q)}$  and using Lemma 4.6, we obtain

$$(56) \quad \sum_{k=1}^q \lambda_k \beta_i^{p_k} = \sup_{S \in \Sigma_{i-1}} \min_{v \in S^\perp, \|v\|_{L^2(\mu)}=1} \langle v, \mathcal{L}^{(q)} v \rangle_{L^2(\mu)} \leq \liminf_{n \rightarrow \infty} \beta_{n,i}.$$

*Proof of the upper bound.* We now derive the corresponding upper bound

$$(57) \quad \limsup_{n \rightarrow \infty} \beta_{n,i} \leq \sum_{k=1}^q \lambda_k \beta_i^{p_k}.$$

We define  $S_n \in \Sigma_{n,i-1}$  to be the span of the orthonormal set  $(\psi_{n,1}, \dots, \psi_{n,i-1})$ . By [16, Max-min theorem], we have

$$\beta_{n,i} = \min_{v \in S_n^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)}$$

and, similarly to the above, without loss of generality, let us assume that  $\limsup_{n \rightarrow \infty} \beta_{n,i} = \lim_{n \rightarrow \infty} \beta_{n,i}$ .

By the induction hypothesis, for each  $1 \leq \ell \leq i-1$ , we have the convergence of eigenvalues and hence

$$\lim_{n \rightarrow \infty} \beta_{n,\ell} = \lim_{n \rightarrow \infty} \langle \mathcal{L}_n^{(q)} \psi_{n,\ell}, \psi_{n,\ell} \rangle_{L^2(\mu_n)} = \beta_\ell < \infty.$$

This uniform boundedness implies that  $\sup_n \langle \mathcal{L}_n^{(q)} \psi_{n,\ell}, \psi_{n,\ell} \rangle_{L^2(\mu_n)} < \infty$  for  $1 \leq \ell \leq i-1$ . We use the same compactness argument as above and a diagonal argument to obtain subsequences - which, to lighten notation, we do not relabel -  $(\mu_n, \psi_{n,\ell})$  converging to  $(\mu, h_\ell)$  in  $TL^2$  for some  $h_\ell \in L^2(\mu)$ . Moreover, by [37, Proposition 2.6] and recalling that  $\langle \psi_{n,\ell}, \psi_{n,j} \rangle_{L^2(\mu_n)} = 0$ , we have orthonormality in the limit

$$\langle h_\ell, h_j \rangle_{L^2(\mu)} = \lim_{n \rightarrow \infty} \langle \psi_{n,\ell}, \psi_{n,j} \rangle_{L^2(\mu_n)} = 0$$

for  $\ell \neq j$  as well as

$$\|h_\ell\|_\rho = \lim_{n \rightarrow \infty} \|\psi_{n,\ell}\|_{\mu_n} = 1$$

for  $1 \leq \ell \leq i-1$ . Let  $S$  be the set spanned by  $\{h_1, \dots, h_{i-1}\}$ . In particular, the above implies that  $S \in \Sigma_{k-1}$ . We also consider  $w \in S^\perp$  such that  $\|w\|_{L^2(\mu)} = 1$  and

$$(58) \quad \langle w, \mathcal{L}^{(q)} w \rangle_{L^2(\mu)} = \min_{v \in S^\perp, \|v\|_{L^2(\mu)}=1} \langle w, \mathcal{L}^{(q)} w \rangle_{L^2(\mu)} \leq \sum_{k=1}^q \lambda_k \beta_i^{p_k},$$

where the last inequality follows from the Courant-Fisher characterization for  $\mathcal{L}^{(q)}$  and Lemma 4.6.

By the lim sup-inequality in [87, Theorem 3.5], we obtain  $w_n \in L^2(\mu_n)$  such that  $(\mu_n, w_n) \rightarrow (\mu, w)$  in  $TL^2(\Omega)$  and  $\limsup_{n \rightarrow \infty} \langle \mathcal{L}_n^{(q)} w_n, w_n \rangle_{L^2(\mu_n)} \leq \langle w, \mathcal{L}^{(q)} w \rangle_{L^2(\mu)}$ . Let us define the projection of  $w_n$  onto the orthogonal complement of  $S_n$  as

$$\tilde{w}_n := w_n - \sum_{\ell=1}^{i-1} \langle w_n, \psi_{n,\ell} \rangle_{L^2(\mu_n)} \psi_{n,\ell}.$$

By construction,  $\tilde{w}_n \in S_n^\perp$ . Moreover, from [37, Proposition 2.6], we have  $\langle w_n, \psi_{n,\ell} \rangle_{L^2(\mu_n)} \rightarrow \langle w, h_\ell \rangle_{L^2(\mu)} = 0$  (since  $w \in S^\perp$ ) as  $n \rightarrow \infty$  for all  $1 \leq \ell \leq i-1$ , and hence, it is straight-forward to check that  $(\mu_n, \tilde{w}_n) \rightarrow (\mu, w)$  in  $TL^2(\Omega)$ .

We next compute the energy of  $\tilde{w}_n$ :

$$\begin{aligned} \langle \mathcal{L}_n^{(q)} \tilde{w}_n, \tilde{w}_n \rangle_{L^2(\mu_n)} &= \left\langle \mathcal{L}_n^{(q)} \left( w_n - \sum_{\ell=1}^{i-1} \langle w_n, \psi_{n,\ell} \rangle_{L^2(\mu_n)} \psi_{n,\ell} \right), w_n - \sum_{m=1}^{i-1} \langle w_n, \psi_{n,m} \rangle_{L^2(\mu_n)} \psi_{n,m} \right\rangle_{L^2(\mu_n)} \\ &= \left\langle \mathcal{L}_n^{(q)} w_n - \sum_{\ell=1}^{i-1} \beta_{n,\ell} \langle w_n, \psi_{n,\ell} \rangle_{L^2(\mu_n)} \psi_{n,\ell}, w_n - \sum_{m=1}^{i-1} \langle w_n, \psi_{n,m} \rangle_{L^2(\mu_n)} \psi_{n,m} \right\rangle_{L^2(\mu_n)} \\ &= \langle \mathcal{L}_n^{(q)} w_n, w_n \rangle_{L^2(\mu_n)} - 2 \sum_{\ell=1}^{i-1} \beta_{n,\ell} \langle w_n, \psi_{n,\ell} \rangle_{L^2(\mu_n)}^2 + \sum_{\ell=1}^{i-1} \beta_{n,\ell} \langle w_n, \psi_{n,\ell} \rangle_{L^2(\mu_n)}^2 \\ &= \langle \mathcal{L}_n^{(q)} w_n, w_n \rangle_{L^2(\mu_n)} - \sum_{\ell=1}^{i-1} \beta_{n,\ell} \langle w_n, \psi_{n,\ell} \rangle_{L^2(\mu_n)}^2. \end{aligned}$$

This implies

$$\limsup_{n \rightarrow \infty} \langle \mathcal{L}_n^{(q)} \tilde{w}_n, \tilde{w}_n \rangle_{L^2(\mu_n)} \leq \limsup_{n \rightarrow \infty} \langle \mathcal{L}_n^{(q)} w_n, w_n \rangle_{L^2(\mu_n)} - \sum_{\ell=1}^{i-1} \beta_{n,\ell} \langle w_n, \psi_{n,\ell} \rangle_{L^2(\mu_n)}^2$$

$$(59) \quad \leq \langle w, \mathcal{L}^{(q)} w \rangle_{L^2(\mu)}$$

where we used the lim sup-inequality of [87, Theorem 3.5] and the fact that  $\beta_{n,\ell} \langle w_n, \psi_{n,\ell} \rangle_{L^2(\mu_n)}^2 \geq 0$  for (59).

Since  $(\mu_n, \tilde{w}_n) \rightarrow (\mu, w)$  in  $\text{TL}^2(\Omega)$  and  $\|w\|_{L^2(\mu)} = 1$ , [37, Proposition 2.6] implies  $\lim_{n \rightarrow \infty} \|\tilde{w}_n\|_{L^2(\mu_n)} = 1$  and we can thus define

$$\bar{w}_n := \frac{\tilde{w}_n}{\|\tilde{w}_n\|_{L^2(\mu_n)}}.$$

We conclude by estimating as follows:

$$(60) \quad \lim_{n \rightarrow \infty} \beta_{n,i} = \lim_{n \rightarrow \infty} \min_{v \in S_n^\perp, \|v\|_{L^2(\mu_n)}=1} \langle \mathcal{L}_n^{(q)} v, v \rangle_{L^2(\mu_n)} \\ \leq \limsup_{n \rightarrow \infty} \langle \mathcal{L}_n^{(q)} \bar{w}_n, \bar{w}_n \rangle_{L^2(\mu_n)}$$

$$(61) \quad \leq \langle w, \mathcal{L}^{(q)} w \rangle_{L^2(\mu)}$$

$$(62) \quad \leq \sum_{k=1}^q \lambda_k \beta_i^{p_k}$$

where we used the fact that  $\|\bar{w}_n\|_{L^2(\mu_n)} = 1$  and  $\bar{w}_n \in S_n^\perp$  for (60), (59) and the fact that  $\lim_{n \rightarrow \infty} \|\tilde{w}_n\|_{L^2(\mu_n)} = 1$  for (61), as well as the facts that  $w \in S^\perp$  and  $\|w\|_{L^2(\mu)}$ , the Courant-Fisher characterization and Lemma 4.6 for (62). This proves (57).

By combining (56) and (57), we get the convergence of eigenvalues. We now consider the convergence of eigenfunctions and proceed similarly by induction.

Before starting, we introduce some additional notation. We denote the ordered eigenvalues of  $\mathcal{L}^{(q)}$  by  $\gamma_i$  (which are equal to  $\sum_{k=1}^q \lambda_k \beta_i^{p_k}$  by Lemma 4.6). We then write  $\bar{\gamma}_i$  for the distinct eigenvalues. Furthermore, for each  $i \in \mathbb{N}$ , let  $s(i)$  denote the multiplicity of the eigenvalue  $\bar{\gamma}_i$ , and let  $\hat{i} \in \mathbb{N}$  be such that

$$\bar{\gamma}_i = \gamma_{i+1} = \dots = \gamma_{i+s(i)}.$$

We define  $E_i$  as the eigenspace of  $\mathcal{L}^{(q)}$  in  $L^2(\mu)$  corresponding to  $\bar{\gamma}_i$ . For  $n$  sufficiently large, let  $E_{n,i} \subset \mathbb{R}^n$  be the subspace spanned by the eigenvectors of  $\mathcal{L}_n^{(q)}$  associated with the eigenvalues  $\beta_{n,\hat{i}+1}, \dots, \beta_{n,\hat{i}+s(i)}$ . Due to the eigenvalue convergence results derived above, we have:

$$(63) \quad \lim_{n \rightarrow \infty} \dim(E_{n,i}) = \dim(E_i) = s(i).$$

We denote by  $\text{Proj}_i : L^2(\mu) \mapsto L^2(\mu)$  the orthogonal projection (with respect to the inner product  $\langle \cdot, \cdot \rangle_{L^2(\mu)}$ ) onto  $E_i$ . Analogously, for all sufficiently large  $n$ , we denote by  $\text{Proj}_{n,i} : L^2(\mu_n) \mapsto L^2(\mu_n)$  the orthogonal projection (with respect to the inner product  $\langle \cdot, \cdot \rangle_{L^2(\mu_n)}$ ) onto the subspace spanned by  $E_{n,i}$ .

The following induction will prove that not only eigenfunctions converge, but also the projections, i.e. if  $(\mu_n, v_n) \rightarrow (\mu, v)$  in  $\text{TL}^2(\Omega)$ , then  $\text{Proj}_{n,i}(v_n) \rightarrow \text{Proj}_i(v)$  in  $\text{TL}^2(\Omega)$ .

**Base case  $i = 1$ .** We covered the convergence of  $\psi_{n,1}$  to  $\psi_1$  in  $\text{TL}^2(\Omega)$  in the base case of the convergence of eigenvalues. Regarding the projections, assume that  $(\mu_n, v_n) \rightarrow (\mu, v)$  in  $\text{TL}^2(\Omega)$ . Since by Assumption S.1  $\Omega$  is connected, the first eigenvalue  $\bar{\gamma}_1 = 0$  is simple, and  $\text{Proj}_1(v)$  corresponds to the constant function equal to the mean of  $v$  with respect to  $\mu$ , that is,  $\text{Proj}_1(v) = \langle v, \mathbb{1} \rangle_{L^2(\mu)}$ . Similarly, convergence of eigenvalue multiplicities (63) implies that for  $n$  large enough,  $E_{n,1}$  is one-dimensional. In this case,  $\text{Proj}_{n,1}(v_n)$  is the constant vector equal to  $\langle v_n, \mathbb{1} \rangle_{L^2(\mu_n)}$ . By [37, Proposition 2.6], we have

$$\lim_{n \rightarrow \infty} \langle v_n, \mathbb{1} \rangle_{L^2(\mu_n)} = \langle v, \mathbb{1} \rangle_{L^2(\mu)},$$

establishing the convergence of the projections.

**Induction step.** Now, suppose that  $\psi_{n,\ell} \rightarrow \psi_\ell$  in  $\text{TL}^2(\Omega)$  and that  $\text{Proj}_{n,\ell}$  converges to  $\text{Proj}_\ell$  for all  $\ell \leq i-1$ . Let  $j \in \{\hat{i}+1, \dots, \hat{i}+s(i)\}$  and consider  $\psi_{n,j}$ . From the convergence of eigenvalues, we have

$$\lim_{n \rightarrow \infty} \langle \mathcal{L}_n^{(q)} \psi_{n,j}, \psi_{n,j} \rangle_{L^2(\mu_n)} = \lim_{n \rightarrow \infty} \beta_{n,j} = \gamma_j < \infty.$$

In particular,  $\sup_n \langle \mathcal{L}_n^{(q)} \psi_{n,j}, \psi_{n,j} \rangle_{L^2(\mu_n)} < \infty$  and we can apply the same compactness result as previously, to obtain a subsequence  $(\mu_{n_m}, \psi_{n_m,j}) \rightarrow (\mu, h_j)$  for some  $h_j \in L^2(\mu)$ .

We note that  $\text{Proj}_{n,\ell}(\psi_{n,j}) = 0$  for all  $1 \leq \ell \leq i-1$  (since  $\psi_{n,j}$  is associated with the eigenvalue  $\beta_{n,i}$ ) and therefore, by the induction hypothesis,  $\text{Proj}_\ell(h_j) = 0$  for all  $1 \leq \ell \leq i-1$ . This allows us to deduce that (using the spectral decomposition of  $\mathcal{L}^{(q)}$ )

$$\langle \mathcal{L}^{(q)} h_j, h_j \rangle_{L^2(\mu)} = \sum_{r=i}^{\infty} \bar{\gamma}_r \|\text{Proj}_r(h_j)\|_{L^2(\mu)}^2 \geq \bar{\gamma}_i \sum_{r=i}^{\infty} \|\text{Proj}_r(h_j)\|_{L^2(\mu)}^2 = \bar{\gamma}_i \|h_j\|_{L^2(\mu)}^2.$$

By [37, Proposition 2.6] and the fact that  $\|\psi_{n,j}\|_{L^2(\mu_n)} = 1$ , we also have  $\|h_j\|_{L^2(\mu)} = 1$ , implying that

$$(64) \quad \langle \mathcal{L}^{(q)} h_j, h_j \rangle_{L^2(\mu)} \geq \bar{\gamma}_i.$$

By using the convergence of eigenvalues, the lim inf-inequality of [87, Theorem 3.5] and (64), we obtain

$$\bar{\gamma}_i = \gamma_j = \lim_{n \rightarrow \infty} \beta_{n,j} = \liminf_{n \rightarrow \infty} \langle \mathcal{L}_n^{(q)} \psi_{n,j}, \psi_{n,j} \rangle_{L^2(\mu_n)} \geq \langle \mathcal{L}^{(q)} h_j, h_j \rangle_{L^2(\mu)} \geq \bar{\gamma}_i.$$

This implies that  $\langle \mathcal{L}^{(q)} h_j, h_j \rangle_{L^2(\mu)} = \bar{\gamma}_i$  and (64) also allows us to deduce that  $\text{Proj}_r(h_j) = 0$  for all  $r \neq i$ . Indeed, suppose that  $\|\text{Proj}_r(h_j)\|_{L^2(\mu)} > 0$  for  $r > i$ . Then (following the same spectral decomposition preceding (64)), we have  $\langle \mathcal{L}^{(q)} h_j, h_j \rangle_{L^2(\mu)} > \bar{\gamma}_i$  which is a contradiction. We conclude that  $h_j$  is an eigenvector of  $\mathcal{L}^{(q)}$  with eigenvalue  $\bar{\gamma}_i$ , establishing the convergence of eigenvectors.

It only remains to prove the convergence of  $\text{Proj}_{n,i}$  to  $\text{Proj}_i$ . Consider  $(\mu_n, w_n) \rightarrow (\mu, w)$  in  $\text{TL}^2(\Omega)$ . According to (63), for sufficiently large  $n$ ,  $\dim(E_{n,i})$  equals  $s(i)$ . We can therefore choose an orthonormal basis  $\{v_{n,1}, \dots, v_{n,s(i)}\}$  of  $E_{n,i}$  with respect to the inner product  $\langle \cdot, \cdot \rangle_{L^2(\mu_n)}$ , where each  $v_{n,j}$  is an eigenvector of  $\mathcal{L}_n^{(q)}$  corresponding to the eigenvalue  $\beta_{n,\hat{i}+j}$ .

Similarly to the above, for each  $j = 1, \dots, s(i)$ , the sequence  $\{v_{n,j}\}_{n \in \mathbb{N}}$  is precompact in  $\text{TL}^2$  and - without relabeling the subsequences - we may assume that  $(\mu_n, v_{n,j}) \rightarrow (\mu, v_j)$  in  $\text{TL}^2(\Omega)$  for some  $v_j \in L^2(\mu)$ . From [37, Proposition 2.6], it follows that each  $v_j$  satisfies  $\|v_j\|_{L^2(\mu)} = 1$ , and that the family  $\{v_1, \dots, v_{s(i)}\}$  is orthonormal with respect to  $\langle \cdot, \cdot \rangle_{L^2(\mu)}$ . Moreover, by the convergence of eigenvectors, each  $v_j$  lies in the limiting eigenspace  $E_i$ , so that  $\{v_1, \dots, v_{s(i)}\}$  forms an orthonormal basis for  $E_i$ . Hence, the projection  $\text{Proj}_i$  can be written for  $v \in L^2(\mu)$  as

$$\text{Proj}_i(v) = \sum_{j=1}^{s(i)} \langle v, v_j \rangle_{L^2(\mu)} v_j.$$

On the discrete side, for all large enough  $n$  and  $v_n \in L^2(\mu_n)$ , we have

$$\text{Proj}_{n,i}(v_n) = \sum_{j=1}^{s(i)} \langle v_n, v_{n,j} \rangle_{L^2(\mu_n)} v_{n,j}.$$

Now, since  $(\mu_n, w_n) \rightarrow (\mu, w)$  and  $(\mu_n, v_{n,j}) \rightarrow (\mu, v_j)$  in  $\text{TL}^2(\Omega)$ , we apply [37, Proposition 2.6] to conclude.  $\square$

**Corollary 4.8** ( $\Gamma$ -convergence of quadratic forms). *Assume that **S.I**, **M.I**, **M.2**, **W.1**, and **D.1** hold. Let  $q \geq 1$ ,  $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$  with  $p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$ . Assume that  $\rho \in C^\infty$ . Assume that  $\varepsilon_n^{(q)}$  satisfies **L.I**. Then,  $\mathbb{P}$ -a.e., for every  $s > 0$ , the following holds:*

1.  $\langle v_n, \left(\mathcal{L}_n^{(q)}\right)^s v_n \rangle_{L^2(\mu_n)}$   $\Gamma$ -converges to  $\langle v, \left(\mathcal{L}^{(q)}\right)^s v \rangle_{L^2(\mu)}$ ;

2. If a sequence satisfies  $\sup_n \max\{\langle v_n, (\mathcal{L}_n^{(q)})^s v_n \rangle_{L^2(\mu_n)}, \|v_n\|_{L^2(\mu_n)}\} \leq C$ , then there exists a converging subsequence in  $TL^2(\Omega)$ .

*Proof.* We want to proceed as in the proof of [23, Theorem 2] where the analogous statement is proven for  $\Delta_{n,\varepsilon_n}^s$  (see also the proof of Proposition 4.5 for a similar argument). In particular, the authors mainly rely on the fact that the eigenpairs of  $\Delta_{n,\varepsilon_n}$  converge to the eigenpairs of  $\Delta_\rho$ . In our case, by Proposition 4.7, the eigenpairs of  $\mathcal{L}_n^{(q)}$  converge to eigenpairs of  $\mathcal{L}^{(q)}$  and we can therefore apply the same argument to conclude.  $\square$

**Proposition 4.9** (Bounded energies). *Assume that S.1, M.1, M.2, W.1 and D.1 hold. Let  $q \geq 1$ ,  $P = \{p_k\}_{k=1}^q \subseteq \mathbb{R}$  with  $p_1 \leq \dots \leq p_q$  and  $E_n = \{\varepsilon_n^{(k)}\}_{k=1}^q$  with  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$ . Assume that  $\rho \in C^\infty$  and that  $\varepsilon_n^{(q)}$  satisfies*

$$(65) \quad \lim_{n \rightarrow \infty} \frac{\log(n)}{n (\varepsilon_n^{(q)})^{d+4p_q}} = 0.$$

For a continuous function  $v$ , let  $v_n$  denote its restriction to  $\Omega_n$ . For any  $k \in \mathbb{N}$  and  $u \in C^\infty(\Omega)$ ,  $\mathbb{P}$ -a.e., there exists a constant  $C(k, u) > 0$  such that

$$\sup_n \langle v_n, (\mathcal{L}_n^{(q)})^{(2k)} v_n \rangle_{L^2(\mu_n)} \leq C(k, u).$$

*Proof.* We want to proceed as in the proof of [89, Lemma 4.19] where the results was shown for the energy  $\langle v_n, \Delta_{n,\varepsilon_n} v_n \rangle_{L^2(\mu_n)}$ . In particular, the proof relies on the following elements:

1.  $\Delta_{n,\varepsilon_n}$  and  $\Delta_\rho$  are self-adjoint and positive semi-definite;
2.  $\langle v_n, \Delta_{n,\varepsilon_n}^s v_n \rangle_{L^2(\mu_n)}$   $\Gamma$ -converges to  $\langle v_n, \Delta_\rho^s v_n \rangle_{L^2(\mu_n)}$ ;
3. There exists a constant  $C(u)$  such that  $\|\Delta_\rho(u) - \Delta_{n,\varepsilon_n}(u)\|_{L^2(\mu_n)} \leq C(u)\varepsilon_n \rightarrow 0$  [35, Theorem 2.8].

For our energy,  $\langle v_n, \mathcal{L}_n^{(q)} v_n \rangle_{L^2(\mu_n)}$ , we have:

1.  $\mathcal{L}_n^{(q)}$  and  $\mathcal{L}^{(q)}$  are positive semi-definite and self-adjoint as shown in Proposition 4.7.
2. the fact that  $\langle v_n, (\mathcal{L}_n^{(q)})^s v_n \rangle_{L^2(\mu_n)}$   $\Gamma$ -converges to  $\langle v, (\mathcal{L}^{(q)})^s v \rangle_{L^2(\mu)}$  is shown in Corollary 4.8.
3. the fact that  $\|\mathcal{L}_n^{(q)}(u) - \mathcal{L}^{(q)}(u)\|_{L^2(\mu)} \leq C(u) \sum_{k=1}^q \lambda_k \varepsilon_n^{(k)} \rightarrow 0$ . Specifically, let  $E_k$  be the set such that [35, Theorem 2.8] holds for  $\varepsilon_n^{(k)}$ : we can apply the latter result since the assumptions that  $p_1 \leq \dots \leq p_q$  and  $\varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$  imply that (65) holds for any  $1 \leq k \leq q$ . We know from the proof of Proposition 4.4 that, for any  $\alpha > 1$ , there exists  $0 < c < C$  and  $\varepsilon_0 > 0$  such that  $\mathbb{P}(\cap_{k=1}^q E_k) \geq 1 - Cn^{-\alpha} - Cne^{-cn(\varepsilon_n^{(q)})^{d+4p_q}}$  as long as  $\varepsilon_0 \geq \varepsilon_n^{(1)} > \dots > \varepsilon_n^{(q)}$ . On this intersection, we have

$$\begin{aligned} \|\mathcal{L}_n^{(q)} u - \mathcal{L}^{(q)} u\|_{L^2(\mu)} &\leq \sum_{k=1}^q \lambda_k \left\| \left( \Delta_{n,\varepsilon_n^{(k)}}^{p_k} - \Delta_\rho^{p_k} \right) u \right\|_{L^2(\mu)} \\ &\leq C \sum_{k=1}^q \lambda_k \varepsilon_n^{(k)} \left( \|u\|_{C^{2p_k+1}(\Omega)} + 1 \right) \\ &= C(u) \sum_{k=1}^q \lambda_k \varepsilon_n^{(k)} \end{aligned}$$

where we used [32, Theorem 2.8] for the inequality. The last term tends to 0 and, by applying the Borel-Cantelli lemma with (65), we can show that this convergence holds  $\mathbb{P}$ -a.e..

We therefore apply the same argument as in [89, Lemma 4.19] to deduce the claim.  $\square$

*Proof of Theorem 3.3.* We are going to proceed as in the proof of Proposition 4.5 where the same result is proven for the truncated energy of a single Laplacian matrix  $\Delta_{n,\varepsilon_n}^s$ . If we replace the latter by  $\mathcal{L}_n^{(q)}$ , [89, Lemma 4.19] by Proposition 4.9, the convergence of eigenpairs by Proposition 4.7 and [89, Proposition 4.21] by Corollary 4.8 the same proof applies.  $\square$

### 4.3 Non-geometric setting

*Proof of Proposition 3.4.* We start by showing that the set of matrices

$$\mathcal{M} = \left\{ M \in \mathbb{R} \mid (M)_{ii} = - \sum_{j \neq i} (M)_{ij} \right\}$$

is closed under matrix product, and addition and multiplication by scalars. Closures under addition and multiplication by scalars are straight-forward to check. Let  $P, Q \in \mathcal{M}$  and consider

$$\sum_{j \neq i} (PQ)_{ij} = \sum_{j \neq i} \sum_{k=1}^n (P)_{ik} (Q)_{kj} = \sum_{k=1}^n (P)_{ik} \sum_{j \neq i} (Q)_{kj} = - \sum_{k=1}^n (P)_{ik} (Q)_{ki} = -(PQ)_{ii}$$

since, by assumption on  $Q$ ,  $(Q)_{kk} = - \sum_{j \neq k} (Q)_{kj}$  implying that  $(Q)_{kk} = - \sum_{j \neq i} (Q)_{kj} - (Q)_{ki} + (Q)_{kk}$  or  $(Q)_{ki} = - \sum_{j \neq i} (Q)_{kj}$ . This implies that  $PQ \in \mathcal{M}$ .

Now, by definition, any Laplacian matrix  $L$  is in  $\mathcal{M}$  and, by the above, so is  $L^k$  for any  $k \in \mathbb{N}$ . Furthermore, since  $L$  is symmetric,  $L^k$  is too. This implies that  $\mathcal{L}_{\text{dis}}^{(q)} = \sum_{k=1}^q \lambda_k (L^{(k)})^k$  is symmetric and in  $\mathcal{M}$ .

Let us now define a graph  $\tilde{G} = (V, W)$  where  $V$  is the same set of vertices used to define  $L_n$  (in our case, this corresponds to  $\Omega_n$  but our proof holds for any set of vertices) and  $W$  is the symmetric matrix with entries  $(W)_{ij} = -(\mathcal{L}_{\text{dis}}^{(q)})_{ij}$  for  $i \neq j$  and  $(W)_{ii}$  can be arbitrarily chosen. Then, for the (diagonal) degree matrix  $D$  with entries  $(D)_{ii} = \sum_{j=1}^n (W)_{ij}$ , the Laplacian of  $\tilde{G}$  defined as  $D - W$  is equal to  $\mathcal{L}_{\text{dis}}^{(q)}$ . Finally, since  $\mathcal{L}_{\text{dis}}^{(q)}$  is a sum of positive semi-definite matrices, it is too and, therefore (2) is a quadratic form.  $\square$

## 5 Numerical experiments

We present experiments illustrating HOHL’s flexibility and effectiveness. First, we show it can replace Laplace learning in active learning. Then, we apply HOHL to hypergraph-structured datasets, observing consistent gains over standard baselines.

### 5.1 Active learning

Optimization problems of the form  $\arg \min_v J(v) + \Psi(v, y)$ , where  $J$  is a regularizer and  $\Psi$  enforces label fidelity, admit a Bayesian interpretation. Specifically, with a prior  $\mu_0(v)$  proportional to  $e^{-J(v)}$ , a likelihood  $\mu_1(y|v)$  proportional to  $e^{-\Psi(v, y)}$ , we obtain a posterior  $\mu_2(v|y)$  that is proportional to  $e^{-J(v) - \Psi(v, y)}$  implying that the maximum a posteriori estimator of  $\mu_2$  is the minimizer of  $J(v) + \Psi(v, y)$ . This formulation enables uncertainty quantification and active learning, see [47, 60, 99].

Active learning is an iterative learning paradigm in which the most informative data points to label are selected at each iteration by an acquisition function, rather than passively relying on a fixed labeled dataset. The goal is to achieve high prediction accuracy with as few labeled examples as possible, making it especially valuable in scenarios where labeling is expensive or time-consuming. Within the Bayesian framework, uncertainty estimates derived from the posterior distribution  $\mu_2(v | y)$  can guide this selection process—for instance, by querying points where the predictive variance is high. This uncertainty-aware strategy helps prioritize data that is expected to most improve the model.

In graph-based approaches, the regularizer is often chosen as  $J(v) = \langle v, L^s v \rangle_n$  for some  $s > 0$ , where  $L$  is the graph Laplacian [23, 60, 84, 98]. This choice induces a Gaussian prior over functions, leveraging the fact that  $L$  is a symmetric and positive semi-definite matrix [84]. By Proposition 3.4, an analogous construction is possible on hypergraphs using the operator  $\mathcal{L}_{\text{Dis}}^{(q)}$ , allowing us to define Gaussian priors in the hypergraph setting as well. This introduces higher-order structure into the prior, effectively encoding regularity up to the  $p_q$ -th derivative [87].

We evaluate this approach within an active learning setting, employing uncertainty sampling as the acquisition function [72]. Experiments are conducted on the MNIST [54] and FashionMNIST [91] datasets. Since both datasets can be embedded in metric spaces, we approximate HOHL by (3) and, following standard practice to speed-up computation on large datasets [11], we construct  $k$ -nearest neighbor graphs instead, replacing



the scale sequence  $\varepsilon^{(\ell)}$  in Eq. (3) with neighborhood sizes  $k^{(1)} \geq \dots \geq k^{(q)}$ . Edge weights are defined by  $w_{k^{(\ell)},ij} = \exp\left(-\frac{4\|x_i - x_j\|^2}{d_{k^{(\ell)}}(x_i)^2}\right)$ , where  $d_{k^{(\ell)}}(x_i)$  is the distance from  $x_i$  to its  $k^{(\ell)}$ -th nearest neighbor. We choose the norm  $\|\cdot\|$  to be the cosine/angular distance.

We compare Laplacian and HOHL-based priors across 100 trials. As shown in Figure 3, HOHL priors yield substantial improvements over graph-based priors, particularly at low label rates where higher-order smoothness improves sample efficiency: with only 100 labeled points (i.e., 0.17% of MNIST and 0.20% of FashionMNIST), on MNIST, accuracy improves from approximately 35% to 75% (+40 points), and on FashionMNIST from 35% to 65% (+30 points), highlighting HOHL’s ability to leverage higher-order structure under severe label constraints.

Our results suggest that smoother priors in high-density regions enable more informative sampling in early rounds, which is critical when label budgets are small.

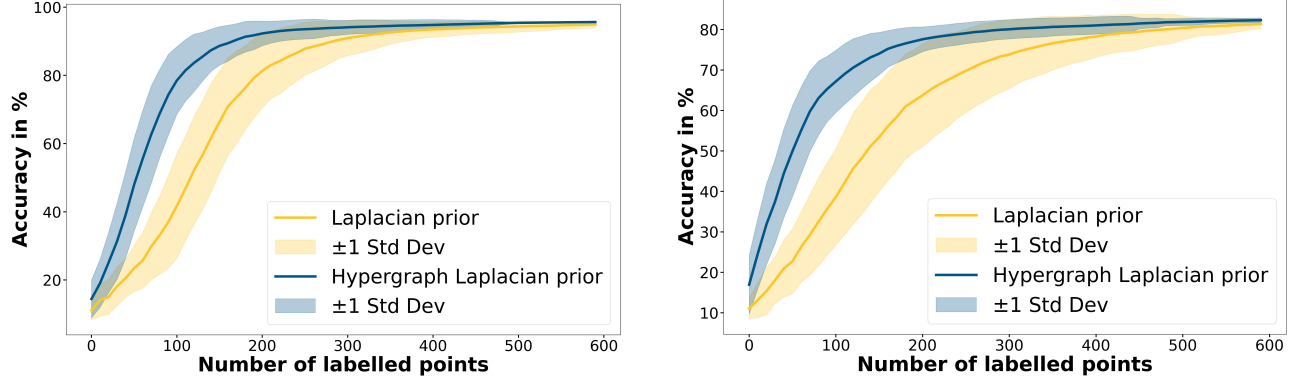


Figure 3: Accuracy in active learning using Laplacian and HOHL priors. We use  $k^{(1)} = 50$ ,  $k^{(2)} = 30$ ,  $\lambda_1 = 1$ ,  $\lambda_2 = 4$ ,  $p_1 = 1$ ,  $p_2 = 2$ . Left: MNIST dataset. Right: fashionMNIST dataset.

## 5.2 HOHL for semi-supervised learning in non-geometric setting

We consider the Zoo [22], Mushroom [22], Cora [58] and Citeseer [71] datasets. The hyperedges are created following the procedure detailed in Section 3.4.3. To ease notation, in this section, we will write  $\mathcal{L}^{(q)}$  instead of  $\mathcal{L}_{\text{Dis}}^{(q)}$ .

Using Algorithm 1, we consider the HOHL energy (2) for semi-supervised learning with

- $1 \leq q \leq 4$ ;
- powers  $p_\ell = \ell$ ;
- regular growth coefficients (RC)  $\lambda_\ell = \ell$  or quickly growing coefficients (QC)  $\lambda_\ell = \ell^2$  (QC).

We compare against Laplace Learning using the clique expansion—chosen over other hypergraph-to-graph reductions for its preservation of the vertex set, see [92]—as well as three non-deep hypergraph methods implemented in [31]: transductive learning from [95], hyperedge-weighted transduction from [30], and dynamic hypergraph learning from [94]. We report mean accuracies and standard deviation in percentages over 100 trials at different labelling rates in Tables 3, 4, 5 and 6. We summarize the terminology used in our experiments in Table 2. Similar experiments have been performed to test HOHL in the geometric setting in [87].

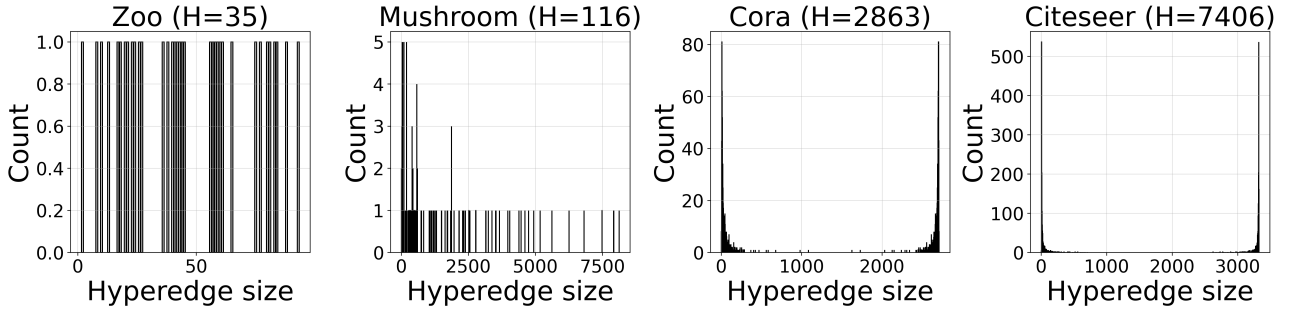


Figure 4: Hyperedge size distributions for all datasets. Zoo and Mushroom exhibit nearly uniform distributions; Cora and Citeseer are bimodal, with both large and small hyperedges.  $H$  denotes the total number of hyperedges in each case.

Term / Abbreviation	Explanation
<b>Aim of experiment</b>	Analysis of HOHL (2) as a function of maximum powers $q$ and coefficients $\lambda_\ell$
$\ell$	Index over scales $1 \leq \ell \leq q$
$q$	Number of Laplacians $1 \leq q \leq 4$
$\lambda_\ell$	Increasing coefficients: $\lambda_\ell = \ell$ or $\lambda_\ell = \ell^2$
$p_\ell$	Increasing powers: $p_\ell = \ell$
<b>RC</b>	$\lambda_\ell = \ell$
<b>QC</b>	$\lambda_\ell = \ell^2$
$\mathcal{L}^{(q)}$	HOHL using Algorithm 1 for $1 \leq q \leq 4$

Table 2: Terminology used in the  $q$ -experiments.

Rate	$\mathcal{L}^{(1)}$	$\mathcal{L}^{(2)}$ RC	$\mathcal{L}^{(2)}$ QC	$\mathcal{L}^{(3)}$ RC	$\mathcal{L}^{(3)}$ QC	$\mathcal{L}^{(4)}$ RC	$\mathcal{L}^{(4)}$ QC
0.05	39.80 (0.00)	42.32 (6.14)	44.69 (7.53)	42.32 (11.57)	33.13 (13.55)	52.33 (8.77)	53.05 (8.03)
0.1	39.78 (0.00)	59.02 (5.52)	62.77 (6.50)	66.91 (13.05)	64.03 (14.72)	74.88 (6.56)	<b>75.35</b> (6.59)
0.2	39.76 (0.00)	75.52 (6.15)	75.88 (5.00)	79.83 (4.28)	77.05 (5.04)	81.95 (3.25)	<b>82.14</b> (2.90)
0.3	39.73 (0.00)	80.56 (1.93)	80.56 (1.64)	83.21 (3.99)	81.05 (4.27)	83.68 (2.82)	<b>83.70</b> (2.81)
0.5	40.38 (0.00)	84.98 (3.22)	85.38 (3.34)	85.06 (2.86)	83.13 (3.24)	<b>85.92</b> (3.49)	<b>85.92</b> (3.43)
0.8	40.91 (0.00)	86.59 (4.49)	<b>87.91</b> (4.10)	87.55 (3.63)	85.68 (4.16)	84.68 (3.86)	84.86 (3.88)

Rate	clique	transductive	weighted transductive	dynamic transductive
0.05	39.80 (0.00)	<b>55.63</b> (3.57)	<b>55.63</b> (3.57)	39.80 (0.00)
0.1	39.78 (0.00)	56.96 (2.02)	56.96 (2.02)	39.78 (0.00)
0.2	39.76 (0.00)	57.37 (1.27)	57.37 (1.27)	39.76 (0.00)
0.3	39.73 (0.00)	58.18 (1.60)	58.18 (1.60)	39.73 (0.00)
0.5	40.38 (0.00)	58.46 (1.83)	58.46 (1.83)	40.38 (0.00)
0.8	40.91 (0.00)	57.50 (2.69)	57.50 (2.69)	40.91 (0.00)

Table 3: Accuracy of various SSL methods on the Zoo dataset. The best-performing method in each row is highlighted in bold.

We observe that HOHL with  $\mathcal{L}^{(q)}$  and  $2 \leq q \leq 4$  consistently either closely matches or achieves higher accuracy than both baseline hypergraph methods and the Laplacian on the clique-expanded graph. This suggests that the skeleton-based segmentation employed by Algorithm 1 succeeds in isolating subgraphs that reflect rel-

Rate	$\mathcal{L}^{(1)}$	$\mathcal{L}^{(2)}$ RC	$\mathcal{L}^{(2)}$ QC	$\mathcal{L}^{(3)}$ RC	$\mathcal{L}^{(3)}$ QC	$\mathcal{L}^{(4)}$ RC	$\mathcal{L}^{(4)}$ QC
0.05	51.79 (0.00)	86.34 (0.81)	86.30 (0.83)	88.70 (1.06)	88.39 (1.19)	63.42 (5.55)	88.00 (1.31)
0.1	51.80 (0.00)	87.22 (0.38)	87.13 (0.38)	88.43 (0.79)	88.45 (0.79)	78.99 (3.17)	88.87 (0.95)
0.2	65.71 (3.76)	88.26 (0.39)	88.34 (0.45)	90.57 (0.69)	90.60 (0.83)	86.92 (1.47)	<b>91.87</b> (0.77)
0.3	84.86 (1.01)	89.20 (0.36)	89.32 (0.28)	92.54 (0.52)	92.45 (0.71)	89.31 (1.05)	<b>93.27</b> (0.45)
0.5	89.74 (0.31)	90.36 (0.49)	90.27 (0.54)	94.22 (0.31)	94.20 (0.44)	89.65 (0.55)	<b>94.27</b> (0.45)
0.8	89.53 (0.63)	91.32 (0.72)	91.29 (0.70)	<b>94.68</b> (0.51)	94.66 (0.44)	90.03 (0.68)	94.66 (0.44)

Rate	clique	transductive	weighted transductive	dynamic transductive
0.05	51.79 (0.00)	<b>90.72</b> (0.67)	90.01 (0.38)	51.79 (0.00)
0.1	51.80 (0.00)	<b>90.80</b> (0.60)	89.96 (0.12)	51.80 (0.00)
0.2	69.72 (3.33)	90.66 (0.34)	90.02 (0.31)	51.80 (0.00)
0.3	85.70 (0.87)	90.65 (0.34)	90.13 (0.27)	51.79 (0.00)
0.5	89.69 (0.35)	90.62 (0.33)	90.24 (0.42)	51.80 (0.00)
0.8	89.73 (0.68)	90.56 (0.64)	90.38 (0.13)	51.78 (0.00)

Table 4: Accuracy of various SSL methods on the Mushroom dataset. The best-performing method in each row is highlighted in bold.

Rate	$\mathcal{L}^{(1)}$	$\mathcal{L}^{(2)}$ RC	$\mathcal{L}^{(2)}$ QC	$\mathcal{L}^{(3)}$ RC	$\mathcal{L}^{(3)}$ QC	$\mathcal{L}^{(4)}$ RC	$\mathcal{L}^{(4)}$ QC
0.05	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.27 (0.12)	<b>30.44</b> (0.49)
0.1	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.29 (0.16)	<b>31.20</b> (0.91)
0.2	30.20 (0.00)	30.20 (0.00)	30.20 (0.00)	30.20 (0.00)	30.20 (0.00)	31.85 (0.67)	<b>34.74</b> (1.49)
0.3	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	35.96 (0.97)	<b>40.15</b> (1.13)
0.5	30.18 (0.00)	30.89 (0.30)	30.89 (0.22)	30.18 (0.00)	30.18 (0.00)	44.39 (1.33)	<b>50.47</b> (1.22)
0.8	30.09 (0.00)	34.86 (0.93)	35.44 (0.92)	30.09 (0.00)	30.09 (0.00)	54.75 (1.49)	<b>60.01</b> (1.42)

Rate	clique	transductive	weighted transductive	dynamic transductive
0.05	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)
0.1	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)
0.2	30.20 (0.00)	30.20 (0.00)	30.20 (0.00)	30.20 (0.00)
0.3	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)	30.19 (0.00)
0.5	30.18 (0.00)	30.18 (0.00)	30.18 (0.00)	30.18 (0.00)
0.8	30.09 (0.00)	30.09 (0.00)	30.09 (0.00)	30.09 (0.00)

Table 5: Accuracy of various SSL methods on the Cora dataset. The best-performing method in each row is highlighted in bold.

Rate	$\mathcal{L}^{(1)}$	$\mathcal{L}^{(2)}$ RC	$\mathcal{L}^{(2)}$ QC	$\mathcal{L}^{(3)}$ RC	$\mathcal{L}^{(3)}$ QC	$\mathcal{L}^{(4)}$ RC	$\mathcal{L}^{(4)}$ QC
0.05	21.06 (0.00)	30.52 (8.39)	31.14 (7.17)	21.13 (0.25)	21.44 (0.92)	33.71 (6.14)	<b>35.14</b> (6.03)
0.1	21.05 (0.00)	40.44 (7.23)	38.04 (10.97)	21.23 (0.24)	21.93 (1.17)	47.14 (4.32)	<b>48.26</b> (3.87)
0.2	21.06 (0.00)	51.13 (3.36)	53.05 (2.65)	22.55 (1.31)	24.26 (2.56)	57.74 (1.58)	<b>57.86</b> (1.40)
0.3	21.06 (0.00)	56.99 (1.83)	56.70 (2.36)	25.15 (1.65)	27.91 (2.12)	<b>61.07</b> (0.99)	60.89 (0.95)
0.5	21.09 (0.00)	62.37 (1.13)	62.52 (1.26)	29.65 (1.15)	34.25 (1.34)	<b>64.08</b> (0.91)	63.66 (0.89)
0.8	21.11 (0.00)	66.04 (1.38)	<b>66.36</b> (1.15)	37.16 (0.75)	43.47 (1.03)	65.63 (1.56)	65.24 (1.57)

Rate	clique	transductive	weighted transductive	dynamic transductive
0.05	21.06 (0.00)	21.06 (0.02)	21.06 (0.00)	21.09 (0.05)
0.1	21.05 (0.00)	21.05 (0.00)	21.05 (0.00)	21.05 (0.00)
0.2	21.06 (0.00)	21.06 (0.00)	21.06 (0.00)	21.06 (0.00)
0.3	21.06 (0.00)	21.06 (0.00)	21.06 (0.00)	21.06 (0.00)
0.5	21.09 (0.00)	21.09 (0.00)	21.09 (0.00)	21.09 (0.00)
0.8	21.11 (0.00)	21.11 (0.00)	21.11 (0.00)	21.11 (0.00)

Table 6: Accuracy of various SSL methods on the Citeseer dataset. The best-performing method in each row is highlighted in bold.

evant structure in the data. In particular, HOHL methods achieve markedly stronger performance on Citeseer and Cora, where conventional hypergraph baselines remain almost flat across all labeling rates — exceeding their accuracy by more than threefold on Citeseer (66.36% for  $\mathcal{L}^{(2)}$  QC vs. 21.11% for baselines at 0.8 label rate) and roughly doubling it on Cora (60.01% for  $\mathcal{L}^{(4)}$  QC vs. 30.09% for baselines at 0.8 label rate).

We also perform an ablation study comparing HOHL with only first-order regularization  $\mathcal{L}^{(1)}$  to the higher-order variant  $\mathcal{L}^{(q)}$  with  $2 \leq q \leq 4$ . The consistent performance gains from adding the higher-order terms suggest that higher-order regularization significantly enhances HOHL’s ability to capture label-relevant structure. The gap between the two versions widens with increasing label rates: on Citeseer, the difference in accuracy grows from 14.08 percentage points at a 0.05 label rate (35.14% for  $\mathcal{L}^{(4)}$  QC vs. 21.06% for  $\mathcal{L}^{(1)}$ ) to 45.25 points at 0.8 (66.36% for  $\mathcal{L}^{(2)}$  QC vs. 21.11% for  $\mathcal{L}^{(1)}$ ); on Zoo, the gain grows from 13.25 points at a 0.05 label rate (53.05% for  $\mathcal{L}^{(4)}$  QC vs. 39.80% for  $\mathcal{L}^{(1)}$ ) to 47.00 points at 0.8 (87.91% for  $\mathcal{L}^{(2)}$  QC vs. 40.91% for  $\mathcal{L}^{(1)}$ ). This effect is strongest when small hyperedges encode local patterns: taking higher powers of their skeleton Laplacians enforces smoothness across these subsets, yielding sharper decision boundaries.

Furthermore, we note that increasing the value of  $\lambda_\ell$ , i.e. comparing RC and QC configurations, can lead to large improvements: 88.00% for  $\mathcal{L}^{(4)}$  QC vs. 63.42% for  $\mathcal{L}^{(4)}$  RC at 0.05 label rate on Mushroom; 50.47% for  $\mathcal{L}^{(4)}$  QC vs. 44.39% for  $\mathcal{L}^{(4)}$  RC at 0.5 label rate on Cora.

Figure 4 shows variation in hyperedge size distribution across datasets which influences how HOHL captures structure across scales.

- In Zoo, the small dataset size increases the chance that early labeled nodes span both fine and coarse hyperedges, enabling HOHL to leverage multiscale structure even at low label rates. In contrast, Mushroom’s larger size makes early labels less likely to touch smaller, more informative hyperedges. HOHL methods thus surpass the transductive baseline only at higher label rates (starting from 0.2), whereas in Zoo they already outperform it at rate 0.1.
- In Cora and Citeseer, the clear size gap between small and large hyperedges creates a strong separation of local and global interactions. As the label rate increases, small hyperedges become more useful, and HOHL’s higher-order regularization captures these patterns. On Citeseer, accuracy improves from 31.14% to 66.36% across label rates 0.05 to 0.8 for  $\mathcal{L}^{(2)}$  QC, while the transductive baseline stays flat at  $\sim 21\%$ .
- The bimodal nature of the hyperedge size distribution in the Cora and Citeseer datasets suggests that an even number of groupings in Algorithm 1 would better align with the data structure. This intuition is supported by our results:  $\mathcal{L}^{(q)}$  with  $q = 2, 4$  consistently outperform  $\mathcal{L}^{(3)}$  ( $\mathcal{L}^{(3)}$  RC and QC remain flat on Cora;  $\mathcal{L}^{(3)}$  RC and QC achieve 37.16% and 43.47% in comparison with 66.36% for  $\mathcal{L}^{(2)}$  QC and

65.63% for  $\mathcal{L}^{(4)}$  RC at 0.8 label rate on Citeseer). In contrast, for datasets like Zoo and Mushroom, where hyperedge sizes are more evenly distributed, the number of groupings appears less critical. In these cases,  $\mathcal{L}^{(3)}$  performs comparably to  $\mathcal{L}^{(q)}$  with  $q = 2, 4$ , confirming that uniform distributions are less sensitive to the choice of segmentation (at 0.8 label rate on Mushroom, we have 94.68% for  $\mathcal{L}^{(3)}$  RC and 94.66 % for  $\mathcal{L}^{(4)}$  QC).

Dataset	$\mathcal{L}^{(1)}$	$\mathcal{L}^{(2)}$ RC	$\mathcal{L}^{(2)}$ QC	$\mathcal{L}^{(3)}$ RC	$\mathcal{L}^{(3)}$ QC	$\mathcal{L}^{(4)}$ RC	$\mathcal{L}^{(4)}$ QC
Zoo	<b>0.00</b> (0.00)	<b>0.00</b> (0.00)	<b>0.00</b> (0.00)	<b>0.00</b> (0.00)	<b>0.00</b> (0.00)	<b>0.00</b> (0.00)	<b>0.00</b> (0.00)
Mushroom	19.91 (0.12)	21.07 (0.04)	21.07 (0.04)	22.01 (0.11)	22.01 (0.11)	22.05 (0.12)	22.05 (0.12)
Cora	<b>2.79</b> (0.05)	2.82 (0.05)	2.82 (0.05)	2.84 (0.01)	2.84 (0.01)	2.86 (0.02)	2.86 (0.02)
Citeseer	<b>4.31</b> (0.04)	4.36 (0.07)	4.36 (0.07)	4.37 (0.01)	4.37 (0.01)	4.38 (0.03)	4.38 (0.03)

Dataset	clique	transductive	weighted transductive	dynamic transductive
Zoo	<b>0.00</b> (0.00)	<b>0.00</b> (0.00)	0.01 (0.00)	0.14 (0.01)
Mushroom	19.72 (0.04)	<b>4.35</b> (0.10)	41.91 (5.14)	301.26 (2.29)
Cora	<b>2.79</b> (0.03)	8.19 (0.16)	83.85 (27.77)	137.15 (1.01)
Citeseer	4.32 (0.05)	33.02 (0.43)	395.14 (224.17)	553.73 (1.30)

Table 7: Computation time in seconds for various SSL methods at label rate 0.1.

Table 7 reports the average time to solve the learning problem at label rate 0.1 (results are similar at all rates), excluding graph or hypergraph construction, which is performed once and reused across experiments. HOHL methods are run with a fixed, untuned configuration and no hyperparameter optimization. By contrast, the last two hypergraph baselines involve iterative solvers and require tuning of regularization parameters, leading to significantly longer runtimes. Despite its simplicity, HOHL consistently achieves strong performance while being quick to compute, underscoring its practical efficiency.

## 6 Conclusion

On the theoretical side, we proved that HOHL is well-posed as a regularizer in the fully supervised setting and established convergence rates between the discrete graph-based approximation and the underlying continuum target function. We further showed that spectrally truncated variants of HOHL remain consistent in the limit, supporting their use in practice.

On the practical side, we demonstrated that HOHL retains the quadratic structure of Laplace learning, making it a viable drop-in replacement within graph-based pipelines. In particular, we integrated HOHL into an active learning framework and observed substantial performance gains in low-label regimes. To generalize HOHL beyond geometric settings, we proposed a multiscale skeleton aggregation algorithm that enables efficient regularization even in the absence of spatial embeddings. Our approach achieves state-of-the-art performance, and we analyzed the impact of HOHL’s parameters in relation to the hyperedge size distribution of the dataset.

Future work includes analyzing HOHL through the lens of reproducing kernel Hilbert space (RKHS) theory, following approaches such as [93], to derive expected error bounds in the semi-supervised setting as a function of the length-scales. Additionally, adaptive skeleton segmentation and parameter selection strategies—e.g., cross-validation, meta-learning, or Bayesian optimization—could further improve robustness. Finally, integrating HOHL into end-to-end differentiable models may enable closer connections to neural architectures, while extending it to dynamic or multilayer hypergraphs opens avenues for application to temporal and multiplex data.

**Acknowledgments** 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”.

## References

- [1] 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.
- [2] Mikhail Belkin and Partha Niyogi. Using manifold stucture for partially labeled classification. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002.
- [3] Mikhail Belkin and Partha Niyogi. Semi-supervised learning on Riemannian manifolds. *Machine Learning*, 56(1):209–239, 2004.
- [4] Mikhail Belkin and Partha Niyogi. Convergence of Laplacian eigenmaps. In *Advances in Neural Information Processing Systems*, 2007.
- [5] Andrea L. Bertozzi, Xiyang Luo, Andrew M. Stuart, and Konstantinos C. Zygalakis. Uncertainty quantification in graph-based classification of high dimensional data. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):568–595, 2018.
- [6] Andrea Braides.  *$\Gamma$ -convergence for Beginners*. Oxford University Press, 2002.
- [7] Leon Bungert, Jeff Calder, Max Mihailescu, Kodjo Houssou, and Amber Yuan. Convergence rates for Poisson learning to a Poisson equation with measure data, 2024.
- [8] 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.
- [9] Jeff Calder. The game theoretic p-Laplacian and semi-supervised learning with few labels. *Nonlinearity*, 32(1):301, dec 2018.
- [10] 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.
- [11] 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.
- [12] Jeff Calder and Nicolás García Trillos. Improved spectral convergence rates for graph Laplacians on  $\varepsilon$ -graphs and  $k - nn$  graphs. *Applied and Computational Harmonic Analysis*, 60:123–175, 2022.
- [13] Jeff Calder and Dejan Slepčev. Properly-weighted graph Laplacian for semi-supervised learning. *Applied Mathematics & Optimization*, 82(3):1111–1159, 2020.
- [14] 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.
- [15] Marco Caroccia, Antonin Chambolle, and Dejan Slepčev. Mumford–Shah functionals on graphs and their asymptotics. *Nonlinearity*, 33(8):3846–3888, jun 2020.
- [16] Isaac Chavel. *Eigenvalues in Riemannian geometry*. Academic Press, Inc, 1984.
- [17] 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.

- [18] 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.
- [19] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.
- [20] Riccardo Cristofori 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.
- [21] Cucker and Smale. Best choices for regularization parameters in learning theory: On the bias—variance problem. *Foundations of Computational Mathematics*, 2(4):413–428, 2002.
- [22] Dheeru Dua and Casey Graff. UCI machine learning repository. <http://archive.ics.uci.edu/ml>, 2017.
- [23] 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.
- [24] 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.
- [25] Ariane Fazeney, 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.
- [26] Lukas Fesser and Melanie Weber. Mitigating over-smoothing and over-squashing using augmentations of forman-ricci curvature. In Soledad Villar and Benjamin Chamberlain, editors, *Proceedings of the Second Learning on Graphs Conference*, volume 231 of *Proceedings of Machine Learning Research*, pages 19:1–19:28. PMLR, 27–30 Nov 2024.
- [27] Mauricio Flores, Jeff Calder, and Gilad Lerman. Analysis and algorithms for  $\ell_p$ -based semi-supervised learning on graphs. *Applied and Computational Harmonic Analysis*, 60:77–122, 2022.
- [28] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738, 2015.
- [29] Charless Fowlkes, Serge Belongie, Fan Chung, and Jitendra Malik. Spectral grouping using the Nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):214–225, 2004.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.

- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] Nicolás García Trillos, Dejan Slepčev, and James Von Brecht. Estimating perimeter using graph cuts. *Advances in Applied Probability*, 49(4):1067–1090, 2017.
- [39] 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.
- [40] Nicolás García Trillos and Dejan Slepčev. On the rate of convergence of empirical measures in  $\infty$ -transportation distance. *Canadian Journal of Mathematics*, 67(6):1358–1383, 2015.
- [41] 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.
- [42] Jhony H. Giraldo, Konstantinos Skianis, Thierry Bouwmans, and Fragkiskos D. Malliaros. On the trade-off between over-smoothing and over-squashing in deep graph neural networks. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM '23*, page 566–576, New York, NY, USA, 2023. Association for Computing Machinery.
- [43] 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.
- [44] Matthias Hein. Uniform convergence of adaptive graph-based regularization. In *Proceedings of the Conference on Learning Theory*, pages 50–64, 2006.
- [45] 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.
- [46] 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.
- [47] Ming Ji and Jiawei Han. A variance minimization criterion to active learning on graphs. In Neil D. Lawrence and Mark Girolami, editors, *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pages 556–564, La Palma, Canary Islands, 21–23 Apr 2012. PMLR.
- [48] Jürgen Jost and Raffaella Mulas. Hypergraph Laplace operators for chemical reaction networks. *Advances in Mathematics*, 351:870–896, 2019.
- [49] Jürgen Jost, Raffaella Mulas, and Dong Zhang. p-Laplace operators for oriented hypergraphs. *Vietnam Journal of Mathematics*, 50(2):323–358, 2022.
- [50] 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.



- [51] Kedar Karhadkar, Pradeep Kr. Banerjee, and Guido Montufar. FoSR: First-order spectral rewiring for addressing oversquashing in GNNs. In *The Eleventh International Conference on Learning Representations*, 2023.
- [52] Anees Kazi, Luca Cosmo, Seyed-Ahmad Ahmadi, Nassir Navab, and Michael M. Bronstein. Differentiable graph module (dgm) for graph convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1606–1617, 2023.
- [53] 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.
- [54] 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.
- [55] Giovanni Leoni. *A First Course in Sobolev Spaces*. Graduate studies in mathematics. American Mathematical Society, 2017.
- [56] 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.
- [57] 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.
- [58] Andrew K. McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3(2):127–163, 2000.
- [59] Ekaterina Merkurjev, Duc Duy Nguyen, and Guo-Wei Wei. Multiscale Laplacian learning. *Applied Intelligence*, 53(12):15727–15746, nov 2022.
- [60] Kevin S. Miller and Andrea L. Bertozzi. Model change active learning in graph-based semi-supervised learning. *Communications on Applied Mathematics and Computation*, 6(2):1270–1298, 2024.
- [61] 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.
- [62] 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.
- [63] Khang Nguyen, Hieu Nong, Vinh Nguyen, Nhat Ho, Stanley Osher, and Tan Nguyen. Revisiting over-smoothing and over-squashing using Ollivier-Ricci curvature. In *Proceedings of the 40th International Conference on Machine Learning, ICML’23*. JMLR.org, 2023.
- [64] Braxton Osting and Todd Harry Reeb. Consistency of Dirichlet partitions. *SIAM Journal on Mathematical Analysis*, 49(5):4251–4274, 2017.
- [65] Bruno Pelletier and Pierre Pudlo. Operator norm convergence of spectral clustering on level sets. *Journal of Machine Learning Research*, 12(12):385–416, 2011.
- [66] Mathew D. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.
- [67] 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.
- [68] Tim Roith and Leon Bungert. Continuum limit of Lipschitz learning on graphs. *Foundations of Computational Mathematics*, pages 1–39, 2022.

- [69] 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.
- [70] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Non-linear Differential Equations and Their Applications*. Birkhäuser Basel, 2015.
- [71] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Gallagher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.
- [72] Burr Settles. *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Springer, Cham, 2012.
- [73] Kehan Shi and Martin Burger. Hypergraph  $p$ -Laplacian equations for data interpolation and semi-supervised learning, 2025.
- [74] Zuoqiang Shi, Stanley Osher, and Wei Zhu. Weighted nonlocal Laplacian on interpolation from sparse data. *Journal of Scientific Computing*, 73(2):1164–1177, 2017.
- [75] Amit Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21:128–134, 2006.
- [76] 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.
- [77] Dejan Slepčev and Matthew Thorpe. Analysis of  $p$ -Laplacian regularization in semisupervised learning. *SIAM Journal on Mathematical Analysis*, 51(3):2085–2120, 2019.
- [78] Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. A survey of signed network mining in social media. *ACM Comput. Surv.*, 49(3), August 2016.
- [79] 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.
- [80] 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.
- [81] Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. In *International Conference on Learning Representations*, 2022.
- [82] Yves van Gennip and Andrea Bertozzi. Gamma-convergence of graph Ginzburg–Landau functionals. *Advances in Differential Equations*, 17(11–12):1115–1180, 2012.
- [83] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer-Verlag Berlin Heidelberg, 2009.
- [84] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 2007.
- [85] Ulrike von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, 36(2):555–586, 2008.
- [86] Xu Wang. Spectral convergence rate of graph Laplacian. *preprint arXiv:1510.08110*, 2015.
- [87] Adrien Weihs, Andrea Bertozzi, and Matthew Thorpe. Analysis of semi-supervised learning on hyper-graphs, 2025.
- [88] 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.

- [89] 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.
- [90] 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.
- [91] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. <https://arxiv.org/abs/1708.07747>, 2017. arXiv:1708.07747.
- [92] 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.
- [93] Tong Zhang and Rie Kubota Ando. Analysis of spectral kernel design based semi-supervised learning. In *Advances in Neural Information Processing Systems*, volume 18, 2005.
- [94] 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.
- [95] 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.
- [96] Dengyong Zhou and Bernhard Schölkopf. Learning from labeled and unlabeled data using random walks. In Carl Edward Rasmussen, Heinrich H. Bühlhoff, Bernhard Schölkopf, and Martin A. Giese, editors, *Pattern Recognition*, pages 237–244, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.
- [97] 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.
- [98] 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.
- [99] Xiaojin Zhu, John Lafferty, and Zoubin Ghahramani. Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions. In *ICML 2003 workshop on the continuum from labeled to unlabeled data in machine learning and data mining*, volume 3, pages 58–65, 2003.