Manifold limit for the training of shallow graph convolutional neural networks

Johanna Tengler*, Christoph Brune*, and José A. Iglesias*

Abstract

We study the discrete-to-continuum consistency of the training of shallow graph convolutional neural networks (GCNNs) on proximity graphs of sampled point clouds under a manifold assumption. Graph convolution is defined spectrally via the graph Laplacian, whose low-frequency spectrum approximates that of the Laplace-Beltrami operator of the underlying smooth manifold, and shallow GCNNs of possibly infinite width are linear functionals on the space of measures on the parameter space. From this functional-analytic perspective, graph signals are seen as spatial discretizations of functions on the manifold, which leads to a natural notion of training data consistent across graph resolutions. To enable convergence results, the continuum parameter space is chosen as a weakly compact product of unit balls, with Sobolev regularity imposed on the output weight and bias, but not on the convolutional parameter. The corresponding discrete parameter spaces inherit the corresponding spectral decay, and are additionally restricted by a frequency cutoff adapted to the informative spectral window of the graph Laplacians. Under these assumptions, we prove Γ-convergence of regularized empirical risk minimization functionals and corresponding convergence of their global minimizers, in the sense of weak convergence of the parameter measures and uniform convergence of the functions over compact sets. This provides a formalization of mesh and sample independence for the training of such networks.

Keywords: graph convolutional neural networks, empirical risk minimization, discrete-to-continuum limit, Γ-convergence, graph Laplacian, Laplace-Beltrami operator

1   Introduction

Across a variety of machine learning scenarios, it is common to assume that data points lie on a smooth manifold, which is commonly referred to as the manifold assumption. One such scenario involves high-dimensional data where the intrinsic dimensionality is believed to be much lower. Another is when data is sampled on points of a physical surface in three-dimensional space, as can be the case in applications of medical imaging, computer vision, and scientific machine learning. In both situations, local extrinsic distances between data points play a central role in determining the relation between them, and using these explicitly provides both reductions of complexity that makes formulating algorithms on such datasets practical, and improved predictions through respecting the intrinsic structure of the data. This role of such distances is apparent in unsupervised clustering algorithms such as Cheeger cuts [14, 44] and in semi-supervised learning [42] where only a small number of data points are labeled, and the eventual predictions are in large part determined by the geometry of the assumed underlying smooth manifold. This underlying local geometry also plays a central role in the case in which

---

*Mathematics of Imaging & AI, Department of Applied Mathematics, University of Twente, the Netherlands (j.tengler@utwente.nl, c.brune@utwente.nl, jose.iglesias@utwente.nl).

the inputs or outputs are assumed to be discretizations of functions on surfaces, since building algorithms depending only on distances between points automatically respects basic translational and rotational invariances that should be satisfied by the physical models that one is attempting to approximate.

Constructing geometric proximity graphs over such samples with an appropriately selected neighborhood scale yields a graph Laplacian whose low-frequency spectrum accurately approximates that of the Laplace-Beltrami operator of the underlying manifold (see e.g. [2], [8], [21]). This fact has been extensively leveraged in semi-supervised learning scenarios in which one seeks to assign one label to each data point, and their consistency when the number of data points tends to infinity [22, 18, 9]. Here, instead we consider problems taking *graph signals* defined on those data points as input. In the first instance, we are concerned with supervised classification of those graph signals assigning a real-valued label, but we expect our techniques to generalize to more complex outputs.

This type of task coupled with the manifold assumption mentioned above naturally leads to the use of geometric machine learning methods (see [7] for a general introduction), which have seen a surge of interest in recent years and in a wide variety of applications. One of the cornerstone architectures in this paradigm is that of graph convolutional neural networks (GCNNs) [28, 20], where convolution is defined spectrally as a weight sharing method that leverages the intrinsic structure of the data. If the graphs and training signals are *consistent*, meaning that they arise as discretizations or samplings of an underlying continuum model, it appears natural to formulate the *consistency of training* in continuum terms as well, in the sense that learning outcomes should remain stable as the discretization is refined. Our approach is motivated by two functional-analytic considerations. First, the Euclidean inner product used in the spectral definition of graph convolution is itself an L^2-structure, and in the context of discretizations, it approximates the canonical L^2-inner product on the manifold. This L^2-structure provides the natural analytic setting for signal and parameter spaces. Second, shallow neural networks admit a dual representation as linear functionals over measures on the parameter space, encompassing both infinite-width limits as well as finite networks. This functional-analytic perspective allows us to analyze shallow GCNNs via weak limits in terms of neuron evaluation functionals or weak-* limits in terms of parameter distributions, and to study their training through the Γ-convergence of empirical risk minimization (ERM) functionals. Within this framework we show our main convergence results, namely that if training data across graph resolutions describe the same phenomenon, that is, the graph signals are discretizations of respective continuum signals, then the discrete training problems are indeed a *consistent discretization* of their continuum counterpart.

Motivated by the Convolution Theorem for Euclidean domains, convolution on graphs and manifolds is defined as multiplication in the spectral domain which gives rise to an L^2-structure of the signal and parameter spaces. Here, the convolution operations $*_n, *$ of signals on graphs and manifolds are defined via the eigendecompositions of the graph Laplacians and the Laplace-Beltrami operator, respectively, where n denotes the resolution. The space of signals on the nodes $\mathcal{M}_n$ of a graph is simply $\mathbb{R}^n$, which naturally extends to the continuum setting, where signals are modeled as functions on a manifold $\mathcal{M}$. A coherent view of this is to define signals as L^2-functions defined on the nodes of the graphs or on the manifold, respectively. In the discrete case, $L^2(\mathcal{M}_n, \mu_n)$ is isomorphic to $\mathbb{R}^n$ equipped with the standard Euclidean inner product, where μ_n are the empirical measures induced from μ. This perspective also applies to the parameter spaces with $\Theta_n \subset (\mathbb{R}^n)^3$ and $\Theta \subset (L^2(\mathcal{M}, \mu))^3$.

Although trained networks are nonlinear in their parameters these parametrizations become linear when expressed by measures over the parameter space (see e.g. [1]). These linear parametrizations are well-defined by the Riesz-Markov representation theorem provided the

parameter spaces are locally compact and such that, for every fixed signal, the neural response maps are continuous in the parameter triple and vanishing at infinity. This is easily satisfied in the finite-dimensional setting by restricting the parameter spaces to closed balls, while in the infinite-dimensional setting the continuity and compactness requirement are competing, which calls for the right choice of topology. Restriction to closed balls with the weak topology on L^2 achieves compactness, while the inner product structure of the neural response maps requires additional H^α regularity on the first and last parameter to retain control over the L^2-norm. This is important for the required continuity of the neural response maps, but also required for the uniform convergence result. Since we want to infer consistency from the spectral approximation this requires imposing this regularity assumption also on the first and last parameter of the discrete parameter spaces, as well as a truncation to the meaningful part of the graph spectra, i.e. those elements spanned by the first $K(n)$ eigenvectors, where $K(n)$ denotes the frequency-cutoff.

The duality framework thus provides a natural foundation for studying approximation, convergence, and generalization in shallow GCNNs. To study the discrete-to-continuum limits, we employ spatial and spectral projections from the continuum to the discrete spaces, denoted by $\mathcal{P}_n, \mathcal{S}_n, \mathcal{S}_{n,\alpha} : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$, along with their respective adjoint operators $\mathcal{P}_n^*, \mathcal{S}_n^*, \mathcal{S}_{n,\alpha}^*$. The spectral projections map onto the $K(n)$-dimensional subspaces of the discrete or continuum L^2 or H^α spaces spanned by their respective eigenbases. Combining those we define a projection between the continuum and discrete parameter spaces

$$\mathcal{Q}_{n,\alpha} : \Theta \to \Theta_n, \quad \theta = (a, b, c) \mapsto \mathcal{Q}_{n,\alpha}\theta \coloneqq (\mathcal{S}_{n,\alpha}a, \mathcal{S}_n b, \mathcal{S}_{n,\alpha}c)$$

and its L^2-adjoint $\mathcal{Q}_{n,\alpha}^*$, which can be seen as a minimal norm extension operator. Then indeed for $n \in \mathbb{N}$, the discrete parameter space Θ_n is isomorphic to the finite-dimensional subspace $\mathcal{Q}_{n,\alpha}^*\Theta_n$ of the constructed continuum parameter space Θ.

We show that then indeed the trained networks corresponding to the discrete setting can be seen as a (spectral) discretization in the linear parametrization of a trained network corresponding to the continuum setting and that the discrete training problems converge to a continuum training problem:

First, we establish uniform convergence of the neural responses to signals that originate from the same signal on the manifold:

$$\psi_n(\mathcal{S}_n u, \mathcal{Q}_{n,\alpha}(\cdot)) \xrightarrow{n\to\infty} \psi(u, \cdot) \text{ in } C((\Theta, w)), \quad \text{for all } u \in L^2(\mathcal{M}, \mu).$$

This ensures consistency of neuron evaluation across resolutions for graph signals whose limit is an underlying manifold signal, in terms of the TL^2 convergence first introduced in [23].

Second, we prove Γ-convergence of the total variation regularized empirical risk minimization (ERM) functionals. Given sampled data from a target functional on the space of manifold signals, the discrete ERM functionals (defined on $\mathcal{M}(\Theta)$) Γ-converge to the continuum ERM functional restricted to a compact ball in $\mathcal{M}(\Theta)$ equipped with the weak-* topology. The discrete ERM functionals are extended to the space $\mathcal{M}(\Theta)$ by assigning the value $+\infty$ to measures that do not correspond to valid projections of measures on the discrete parameter space, that is if $(\mathcal{Q}_{n,\alpha}^*\mathcal{Q}_{n,\alpha})_\#\nu \neq \nu$. As a result, minimizers of the discrete ERM functionals converge weak-* to minimizers of the continuum ERM functional, ensuring consistency of the trained networks.

Organization of the paper. In Section 2 we establish the considered setting, terms and notation of this paper, namely we define shallow GCNNs and formalize the manifold assumption, and the notion of consistency used in this paper. In Section 3 we collect spectral approximation results from [21] and [8], verify their applicability to our setting and state them in a form that is

convenient for the sequel. In Section 4 we define spectral projection maps between the discrete and continuum spaces and prove some preliminary convergence results about shallow GCNNs in order to finally state and prove our main results Theorem 5.2 and Theorem 5.6 in Section 5. In Section 6 we discuss some of the assumptions made in this paper on the underlying manifold, as well as the construction of the parameter spaces.

Related works. There are several works studying transferability, approximation and generalization properties of GCNNs, such as [32, 30, 34, 35, 46, 47].

Transferability compares the outputs of a model on different graphs, for instance on graphs of different sizes, relying on some notion of graphs describing the same underlying structure. In contrast, in the present work, we are interested in the consistency of the training of GCNNs, so in convergence of minimizers of risk objectives corresponding to graphs consistent across resolutions with consistent training data.

The authors in [32] analyze transferability of GCNNs by the means of the error of the filter of a continuum Laplacian applied to a signal with the interpolation of the application of a filter of a graph Laplacian to a discretized signal. They show that this transferability error of the filter is controlled by the transferability error of the Laplacian and a so-called signal consistency error, i.e. the error between the signal and the interpolation of its discretization. While we also consider graphs that are consistent in the sense that they are sampled from the same underlying space, in our setting the graph Laplacians are required to approximate the Laplace-Beltrami spectrally in a low-frequency window, while [32] consider graphs whose action on sampled continuum low-frequency modes approximates the action of the continuum-Laplacian under sampling. In their paper, the space of graph signals is limited to have finite bandwidth. Indeed, the fact that only the low-frequency graph-spectrum reliably approximates that of the underlying continuum space makes restriction to low-frequencies unavoidable, however we impose such restrictions on the parameter space, and adapt the discretization scheme accordingly.

Some works on the manifold convolution setting we treat here are [46] and [47], where the authors first claim a bound on the error of commuting the filter operation with the discretization, and subsequently, for filters where this bound is uniform across signals, they show that this error remains controlled through the application of the nonlinearity. Notably, the discretization considered in [46, 47] again requires the space of input manifold signals to be band-limited. We prove a similar type of bound in Lemma 4.11, which we later use to show consistency of the training, as in Γ-convergence of the empirical risk minimization functionals. Furthermore, in [47] the authors consider manifold filters whose response to high frequencies is almost constant. Rather than imposing assumptions on the manifold filter, we shift the restriction to the discretized filter by truncating its spectral representation to the first $K(n)$ frequencies, where the spectral cutoff $K(n)$ is chosen such that it diverges. However, this requires imposing an assumption on the asymptotic behavior of the spectral gap of the manifold. In particular we provide a quantification of this admissible spectral window in terms resolution and this gap assumption.

There are also a number of works relying on general notions of convergence and limiting objects for graphs applicable beyond the geometric setting. One such framework is that of graphons, as considered for instance in [35], which does not apply in our setting, because graphons arise as limit objects of sequences of dense graphs, whereas the geometric proximity graphs considered here are not dense. The reason is that their number of edges per node is bounded relative to the local sampling density, due to being constructed from a kernel with compact support. Another type of convergence that is applicable to sparser graphs is the notion of action convergence of graphops as used in [30]. However, our models are based on convolutions defined in the spectral domain, and as the authors point out, the relations between spectral definitions and this type of convergence are not clear at present.

Finally, let us also mention that further architectures based on spectral constructions are considered in [33], and generalization results derived from the manifold assumption and upscaling in the number of graph vertices are provided in [45].

Table 1: Notation

d, m	Dimensions of the ambient data space $\mathbb{R}^d$ and of the manifold $\mathcal{M}$.
$\mathcal{M}$	Compact connected smooth manifold without boundary of dimension m embedded in $\mathbb{R}^d$, whose Riemannian metric g is the one inherited from the embedding.
$d_{\mathcal{M}}(x, y)$	The geodesic distance between points $x, y \in \mathcal{M}$.
μ	Probability measure supported on $\mathcal{M}$ describing the data distribution.
Δ	The Laplace-Beltrami operator on $\mathcal{M}$.
λ_k	Eigenvalues of the Laplace-Beltrami operator, arranged such that $\lambda_1 \leq \lambda_2 \leq \ldots \uparrow +\infty$
$E(\lambda)$	Eigenspace of the Laplace-Beltrami operator associated to the eigenvalue λ.
$\gamma_{k,\mu}$	The spectral gap of the eigenvalue λ_k, i.e. the minimal distance of the eigenvalue λ_k to its distinct neighboring eigenvalues.
φ_k	Arbitrary choice of an orthonormal basis of $L^2(\mathcal{M}, \mu)$ consisting of eigenfunctions of the Laplace-Beltrami operator in order of increasing eigenvalues.
$N(\lambda)$	Eigenvalue counting function (see (3.4)).
$\mathcal{M}_n$	Point cloud $\mathcal{M}_n = \{x_1, \ldots x_n\} \subset \mathbb{R}^d$ drawn from i.i.d samples from μ, the nodes of the geometric proximity Graph G_n.
n	Resolution or size of the graph G_n, i.e. number of nodes $\mathcal{M}_n := \{x_1, \ldots x_n\}$ of G_n.
μ_n	Empirical measure of the sample $\mathcal{M}_n$.
ε_n	Upper bound on the geodesic distance $d_{\mathcal{M}}$ between any point in $\mathcal{M}$ and a point of $\mathcal{M}_n$ (see (2.1)).
h_n	Radius, in terms of geodesic distance, of interaction between points so that the edge between them receives a nonzero weight (see (2.4)).
Δ_n	Unnormalized graph Laplacian corresponding to the n-th geometric graph G_n, as defined in (2.5).
$\lambda_k^{[n]}$	Eigenvalues of the graph Laplacian, arranged such that $\lambda_1^{[n]} \leq \ldots \leq \lambda_n^{[n]}$
$K(n)$	A frequency cutoff compatible with multiplicity blocks for the n-th geometric proximity graph, such that up to this index the spectrum of the n-th graph Laplacian is informative about that of the Laplace-Beltrami (see (2.8)).
$\varphi_k^{[n]}, \varphi_k^{(n)}$	An arbitrary choice of eigenfunctions $\varphi_k^{[n]}$ of the graph Laplacian corresponding to the eigenvalues $\lambda_k^{[n]}$ such that they form an orthonormal basis of $L^2(\mathcal{M}_n, \mu_n)$ and a resolution-dependent choice of orthonormal eigenfunctions $\varphi_k^{(n)}$ of the Laplace-Beltrami operator corresponding to the eigenvalues $\lambda_k^{(n)}$, such that the first $K(n)$ of them align well with the eigenfunctions $\varphi_k^{[n]}$ of the graph Laplacian for the graph of size n.
$\phi_k^{[n]}, \phi_k^{(n)}$	A fixed choice of orthonormal eigenfunctions of the graph Laplacian corresponding to the eigenvalues $\lambda_k^{[n]}$ and a fixed resolution-dependent choice of orthonormal eigenfunctions of the Laplace-Beltrami operator corresponding to the eigenvalues $\lambda_k^{(n)}$, such that the first $K(n)$ of them align well with the eigenfunctions $\phi_k^{[n]}$ of the graph Laplacian for the graph of size n.
$\delta(\varepsilon, h, \lambda)$	Auxiliary function for bounding approximation errors between discrete and continuum eigenvalues and eigenfunctions, as defined in (3.3).
δ_n^ϕ	Upper bound on the norm of the difference of interpolated discrete eigenfunction and continuous ones, as defined in (3.14).

Θ_n, Θ	Discrete and continuum parameter space, see (2.18), and (2.17).
α	The order of fractional Sobolev regularity imposed on the parameter spaces.
$\ast_n, \ast$	Convolution operations, defined by multiplication in the Fourier domain of the graph Laplacian corresponding to the graph of size n or the Laplace-Beltrami operator, respectively.
$\sigma, L_\sigma, C_\sigma$	The activation function, its Lipschitz constant, and a linear growth constant for it (see (2.10)).
ψ_n, ψ	Neural response maps as defined in (2.13) and (2.14).
$f_\rho^{[n]}$	Infinitely wide shallow graph convolutional neural network parametrized by the measure ρ on the discrete parameter space Θ_n, as defined in (2.11).
f_ρ	Infinitely wide shallow manifold convolutional neural network parametrized by the measure ρ on the continuum parameter space Θ, as defined in (2.12).
$\mathcal{P}_n, \mathcal{P}_n^*$	Spatial discretization and extension maps defined in (2.19) and (2.20).
I_n	Interpolation operator from [21, p. 840], see (2.22).
$\mathcal{R}_n, C_\mathcal{R}$	Linear discretization operators, uniformly bounded in n by $C_\mathcal{R}$, and compatible with TL^2-convergence (see Definition 4.5).
$\mathcal{S}_{n,\alpha}, \mathcal{S}_{n,\alpha}^*$	Spectral discretization and extension maps that respect the H^α-structure, defined in (4.1) and (4.2).
$\mathcal{Q}_{n,\alpha}, \mathcal{Q}_{n,\alpha}^*$	The parameter discretization operation and its adjoint, constructed from the spatial and spectral discretization and extension operators (4.6) and (4.7).
ℓ	Loss function $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, generally assumed to be continuous (in the first argument).
J_α	The empirical risk minimization functional corresponding to the training of the shallow manifold convolutional neural network, as defined in (5.1).
$\tilde{J}_{n,\alpha}$	The empirical risk minimization functional corresponding to the training of the shallow graph convolutional neural network, where the graph is of resolution n and α is the degree of regularity, as defined in (5.2).
$J_{n,\alpha}$	The lifted empirical risk minimization functional corresponding to the training of the graph of resolution n, as defined in (5.3).
$\mathcal{M}(X)$	The space of finite signed Radon measures on X.
$\mathscr{P}(X)$	The space of Radon probability measures on X.
w, w^*	Weak, weak-* topology.
C	General notation for no-further specified constants, whose value may vary across occurrences, even within one formula.
ω_m	The volume of the unit ball in $\mathbb{R}^m$.
$B_V, B_V(v, r)$	The unit ball in the normed vector space $(V, \lVert \cdot \rVert_V)$; a closed ball of radius r in the normed vector space $(V, \lVert \cdot \rVert_V)$ around $v \in V$.
$B_{\lVert \cdot \rVert_{V_w}}$	The unit ball in the normed vector space $(V, \lVert \cdot \rVert_{V_w})$, used if in particular a different norm than the standard norm on V is used.
$B_{d_V}(v, r)$	A closed ball of radius r in the metric space (V, d_V) around $v \in V$.

2 Setup and notation

A comprehensive list of the notation used in the paper is provided in Table 1. In the remainder of this section, we introduce the framework for our subsequent results.

2.1 Point cloud proximity graphs and the manifold assumption

The manifold assumption refers to the physically motivated idea that data points, which are a priori vectors in $\mathbb{R}^d$, lie on a smooth manifold embedded in $\mathbb{R}^d$, implying an underlying geometric

structure of the data sample that one can attempt to exploit. By connecting data points if their Euclidean distance is below a certain threshold, the neighborhood size, one obtains a neighborhood graph from the point cloud, a spatial discretization of the underlying manifold. However, what is ultimately desired is to induce the geometry of the underlying Riemannian manifold onto the graph. The geometry of a Riemannian manifold is completely determined by its Riemannian metric, which itself can be recovered from the Laplace-Beltrami operator: to see this, one can just notice that its principal symbol, the cometric in (2.6), can be recovered through its action on coordinate functions, see [41, Sec. 2.9]. This suggests that if one has a suitable notion of a graph Laplacian approximating the former in an appropriate sense, then the graphs are also a geometric approximation. For a compact Riemannian manifold, the Laplace–Beltrami operator has a discrete spectrum, and thus a natural notion of approximation between such Laplacians is spectral convergence, which has been studied extensively (e.g. [25, 26, 38, 2, 8, 24, 21]).

To obtain a good approximation of the lower-frequency spectrum, the neighborhood size must shrink appropriately with increasing resolution, i.e. the number of nodes, but also account for the sampling error. This will be made explicit in the following, and elaborated more in Section 3, where we will state the main results, Theorems 4 (or its Corollary 1) and 5, from [21] which establish convergence rates for the convergence of eigenvalues and eigenvectors of the graph Laplacians towards eigenvalues and eigenfunctions of the Laplace-Beltrami operator on the underlying manifold.

Let $\mathcal{M} \subset \mathbb{R}^d$ be an m-dimensional compact connected C^∞ Riemannian submanifold without boundary embedded in $\mathbb{R}^d$, with $d \geq m$.

We denote by $\mu \in \mathscr{P}(\mathbb{R}^d)$ (Radon probability measures in $\mathbb{R}^d$) the normalized m-dimensional Hausdorff measure restricted to $\mathcal{M}$, that is

$$\mu = \frac{1}{\mathcal{H}^m(\mathcal{M})} \mathcal{H}^m \llcorner \mathcal{M}, \quad \text{where } (\mathcal{H}^m \llcorner \mathcal{M})(A) = \mathcal{H}^m(A \cap \mathcal{M}) \text{ for all Borel sets } A \subseteq \mathbb{R}^d,$$

which for functions $f : \mathcal{M} \to \mathbb{R}$ satisfies

$$\int_{\mathcal{M}} f(x) \, \mathrm{d}\mu(x) = \frac{1}{\mathrm{Vol}_{\mathcal{M}}(\mathcal{M})} \int_{\mathcal{M}} f(x) \, \mathrm{d}\,\mathrm{Vol}_{\mathcal{M}}(x),$$

where the right integral is against the volume form on $\mathcal{M}$ arising from the Riemannian metric g induced by the embedding of $\mathcal{M}$ into $\mathbb{R}^d$, since following [40, Prop. 12.6] it holds that

$$\mathrm{Vol}_{\mathcal{M}}(A) = \mathcal{H}^m(A) = (\mathcal{H}^m \llcorner \mathcal{M})(A) \qquad \text{for every Borel set } A \subset \mathcal{M}.$$

From μ we sample an i.i.d. sequence of points $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}^d$, and denote the first n of them and corresponding empirical measure as

$$\mathcal{M}_n := \{x_1, \ldots, x_n\}, \quad \mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}.$$

By the general version in [43] (see also [15, Thm. 11.4.1]) of the Glivenko-Cantelli theorem, almost surely the measures μ_n converge weakly to μ. Moreover, this convergence can be quantified in terms of the ∞-Wasserstein distance

$$\varepsilon_n := d_\infty(\mu, \mu_n) := \inf_{T_{\#}\mu=\mu_n} \mathrm{esssup}_{x\in\mathcal{M}} d_{\mathcal{M}}(x, T(x)), \tag{2.1}$$

where $d_{\mathcal{M}}$ is the Riemannian geodesic distance on $\mathcal{M}$ and the infimum is taken over all Borel measurable maps from $\mathcal{M}$ to $\mathcal{M}$. The bounds

$$\varepsilon_n \leq C \frac{(\log n)^{1/m}}{n^{1/m}} \text{ for } m \geq 3, \quad \varepsilon_n \leq C \frac{(\log n)^{3/4}}{n^{1/2}} \text{ for } m = 2, \tag{2.2}$$

7

hold with very high probability by [21, Thm. 2], which in turn is a generalization to the manifold setting of the classical bounds with the same rate for the uniform distribution on $(0,1)^m$ of [37] and [31].

Remark 2.1. The above-cited results on convergence rates for eigenvalues and eigenfunctions as well as the bounds on ε_n hold also for probability distributions with Lipschitz continuous density bounded uniformly away from zero, which is the framework used in [21]. We believe our analysis can be generalized to this case, but for brevity we restrict ourselves to the case of constant density, since otherwise a number of arguments in Section 4 would become more cumbersome.

Let $\{h_n\}_{n\in\mathbb{N}} \subset \mathbb{R}_{>0}$ be a sequence of neighborhood sizes, such that [21, Assumption 3]

$$(m+5)\varepsilon_n < h_n < \min\left\{1, \frac{i_0}{10}, \frac{1}{\sqrt{mK}}, \frac{R}{\sqrt{27m}}\right\}, \quad \text{and} \quad \frac{\varepsilon_n}{h_n} \xrightarrow{n\to\infty} 0, \tag{2.3}$$

where i_0 is the injectivity radius of the manifold $\mathcal{M}$, K a global upper bound on the absolute value of sectional curvatures of $\mathcal{M}$, and R the reach of $\mathcal{M}$. Given an interaction strength $\eta : [0, +\infty) \to [0, +\infty)$ with support in $[0,1]$ and normalized such that

$$\int_{\mathbb{R}^m} \eta(|x|)\,\mathrm{d}x = 1$$

we define, as in [21], the weights

$$W_n = \left(w_{ij}\right)_{i,j=1}^n, \quad w_{ij} = \frac{1}{nh_n^m}\eta\left(\frac{|x_i - x_j|}{h_n}\right), \tag{2.4}$$

in which $|x_i - x_j|$ denotes the Euclidean distance in $\mathbb{R}^d$. Putting these together, we get an undirected graph $G_n = (\mathcal{M}_n, W_n)$ where we keep the edge set implicit in the weights, that is, two points/vertices x_i, x_j are connected if and only if $w_{ij} > 0$. With the above choice of the neighborhood sizes h_n, this yields a connected graph for every $n \in \mathbb{N}$.

We define the "surface tension" associated to η as

$$\sigma_\eta := \int_{\mathbb{R}^m} |x \cdot e_1|^2\,\eta(|x|)\,\mathrm{d}x.$$

This kind of quantity naturally appears as multiplicative factor in the convergence of functionals defined by double integrals to those involving classical derivatives as started in the pioneering work [3], see also the proof of [23, Lem. 4.2] for an illustrative computation deriving its appearance. Accounting for this factor to aim for convergence to the continuous counterpart, for a function $v : \mathcal{M}_n \to \mathbb{R}$, the unnormalized graph Laplacian $\Delta_n v : \mathcal{M}_n \to \mathbb{R}$ is defined by

$$(\Delta_n v)(x_i) = \frac{2}{\sigma_\eta h_n^2} \sum_{j=1}^n w_{ij}\big(v(x_i) - u(x_j)\big). \tag{2.5}$$

On the continuum level, the Laplace-Beltrami operator on $\mathcal{M}$ is defined for $v \in C^2(\mathcal{M})$ by

$$\Delta v = -\frac{1}{\mathrm{Vol}_{\mathcal{M}}(\mathcal{M})}\,\mathrm{div}_g\,\nabla_g v, \quad \text{so that} \quad \int_{\mathcal{M}} w\Delta v\,\mathrm{dVol}_{\mathcal{M}} = \int_{\mathcal{M}} g(\nabla_g w, \nabla_g v)\,\mathrm{d}\mu,$$

where ∇_g and div_g are the gradient and divergence arising from the metric g on $\mathcal{M}$. In coordinates, this leads to the expression

$$-\frac{1}{\mathrm{Vol}_{\mathcal{M}}(\mathcal{M})\sqrt{\det g}} \sum_{i,j=1}^m \partial_j\left(\sqrt{\det g}\, g^{ij}\,\partial_i f\right). \tag{2.6}$$

8

Denote by $\lambda_k^{[n]}$ the k-th eigenvalue of the n-th graph Laplacian Δ_n, where $0 = \lambda_1^{[n]} \le \lambda_2^{[n]} \le \ldots \lambda_n^{[n]}$ and by $\{\phi_k^{[n]}\}_{k=1}^n$ any of their eigenvectors. Since $\mathcal{M}$ is a compact Riemannian manifold, its Laplace-Beltrami operator has a discrete spectrum $0 = \lambda_1^{[n]} \le \lambda_2^{[n]} \le \ldots \uparrow +\infty$ (see e.g.[29, Thm. 4.3.1]). Accordingly, we denote by λ_k, $k \in \mathbb{N}$ the k-th eigenvalue of the Laplace-Beltrami operator on $\mathcal{M}$ and by $\{\phi_k\}_k$ a choice of corresponding orthonormal eigenfunctions. Moreover, we denote by

$$\gamma_{k,\mu} := \min\left\{|\lambda_k - \lambda_j| \mid j \in \mathbb{N}, \lambda_k \ne \lambda_j\right\} > 0 \tag{2.7}$$

the gap between the k-th eigenvalue and any other distinct one. Note that by definition this quantity is nonzero even in the presence of multiplicity for λ_k.

For two sequences $\{a_n\}_n, \{b_n\}_n$ we denote by $0 < a_n \ll b_n$ that for all $n \in \mathbb{N}$ it is $a_n < b_n$ and $a_n/b_n \xrightarrow{n \to \infty} 0$.

Assumption 2.2. *We assume further on the manifold $\mathcal{M}$ that there is a $\beta \ge 0$, such that $\gamma_{k,\mu}^{-1} = O(k^\beta)$ and define*

$$\beta^* := \inf\left\{\beta \ge 0 \mid \gamma_{k,\mu}^{-1} = O(k^\beta)\right\}.$$

Then let $\{\widetilde{K}(n)\}_{n \in \mathbb{N}}$ be a sequence of $\mathbb{N} \ni \widetilde{K}(n) \le n$, such that

$$0 < \varepsilon_n \ll h_n \ll \widetilde{K}(n)^{-\max\{\beta^* + \frac{m+1}{2m}, 1\}} \ll 1, \quad \text{and} \quad \varepsilon_n h_n^{-1} \ll \widetilde{K}(n)^{-\left(\beta^* + \frac{m-1}{2m}\right)}. \tag{2.8}$$

These assumptions are sufficient to ensure that the continuous eigenvalues and eigenspaces are well-approximated by the discrete ones, as we will see in Section 3.

Remark 2.3. Suppose $\beta^* + \frac{m+1}{2m} \ge 1$, i.e. $\beta^* \ge \frac{2m-m-1}{2m} = \frac{m-1}{2m}$. Hence $2m\beta^* \ge m - 1$ and thus

$$\frac{2m}{2m\beta^* + m + 1} \ge \frac{2m}{4m\beta^* + 2} \ge \frac{1}{2\beta^* + 1}.$$

Therefore, for $0 < h_n < 1$ it holds that

$$h_n^{\frac{m}{2m\beta^* + m + 1}} \le h_n^{\frac{1}{2\beta^* + 1}}.$$

Then for a sequence of neighborhood sizes $\{h_n\}_n \subset (0, 1)$, satisfying (2.3), one could choose $\widetilde{K}(n)$ as

$$\widetilde{K}(n) := \min\{\lfloor h_n^{-\frac{m}{2m\beta^* + m + 1}} \rfloor, n\}$$

since

$$h_n^{-\frac{m}{2m\beta^* + m + 1}} \ge h_n^{-\frac{1}{2\beta^* + 1}} \xrightarrow{n \to \infty} +\infty, \qquad \widetilde{K}(n)^{-\left(\beta^* + \frac{m+1}{2m}\right)} \ge h_n^{1/2} > h_n,$$

and

$$\widetilde{K}(n)^{\left(\beta^* + \frac{m+1}{2m}\right)} h_n \le (h_n^{-\frac{m}{2m\beta^* + m + 1}})^{\left(\beta^* + \frac{m+1}{2m}\right)} h_n \le h_n^{1/2} \xrightarrow{n \to \infty} 0.$$

Thus it holds indeed that $h_n \ll \widetilde{K}(n)^{-\left(\beta^* + \frac{m+1}{2m}\right)}$. If one takes $h_n = \sqrt{\varepsilon_n}$, one obtains

$$\varepsilon_n h_n^{-1} = h_n \ll \widetilde{K}(n)^{-\left(\beta^* + \frac{m+1}{2m}\right)} \le \widetilde{K}(n)^{-\left(\beta^* + \frac{m-1}{2m}\right)},$$

so $\{\widetilde{K}(n)\}_n$ satisfies (2.8). Using (2.2) [21, Thm. 2] and the above, it follows that with high probability for $m \ge 3$,

$$\widetilde{K}(n) \ge C\left(\frac{n}{\log(n)}\right)^{\frac{1}{2(2m\beta^* + m + 1)}} \ge C\left(\frac{n}{\log(n)}\right)^{\frac{1}{4(2m\beta^* + 1)}}.$$

Remark 2.4. Now, suppose $\beta^* + \frac{m+1}{2m} < 1$. Then for a sequence of neighborhood sizes $\{h_n\}_n$, satisfying (2.3), one could choose $\widetilde{K}(n)$ as

$$\widetilde{K}(n) := \min\{\lfloor h_n^{-1/2}\rfloor, n\} \xrightarrow{n\to\infty} +\infty,$$

which also yields $\widetilde{K}(n)^{-\left(\beta^* + \frac{m+1}{2m}\right)} \ge \widetilde{K}(n)^{-1} \ge h_n^{1/2} > h_n$, and using that $\beta^* + \frac{m+1}{2m} < 1$ further we have

$$\widetilde{K}(n)^{\left(\beta^* + \frac{m+1}{2m}\right)} h_n \le h_n^{-\frac{2m\beta^* + m + 1}{4m}} h_n \le h_n^{-\frac{2m(\beta^*+1)}{4m}} h_n \le h_n^{\frac{1-\beta^*}{2}} \le h_n^{\frac{m+1}{4m}} \xrightarrow{n\to\infty} 0.$$

With the same argument as in the other case, taking $h_n = \sqrt{\varepsilon_n}$ yields

$$\varepsilon_n h_n^{-1} = h_n \ll \widetilde{K}(n)^{-\left(\beta^* + \frac{m+1}{2m}\right)} \le \widetilde{K}(n)^{-\left(\beta^* + \frac{m-1}{2m}\right)},$$

so again $\{\widetilde{K}(n)\}_n$ satisfies (2.8). In this case it holds that $2m\beta^* < m - 1$, and thus using (2.2) [21, Thm. 2] it follows that with high probability for $m \ge 3$,

$$\widetilde{K}(n) \ge C\left(\frac{n}{\log(n)}\right)^{\frac{1}{2(2m\beta^* + m + 1)}} \ge C\left(\frac{n}{\log(n)}\right)^{\frac{1}{8m}}.$$

From Remark 2.3 and Remark 2.4 it is clear that for any choice of neighborhood radii $\{h_n\}_{n\in\mathbb{N}} \subset (0,1)$ fulfilling (2.3), there always exists a choice of indices $\{\widetilde{K}(n)\}_{n\in\mathbb{N}} \subseteq \mathbb{N}$, $\widetilde{K}(n) \le n$ for all $n \in \mathbb{N}$ which satisfies the first part of (2.8), and if the neighborhood sizes are also chosen adequately in relation to the ∞-transportation distances ε_n, then also the second part of (2.8) is satisfied. Given one such choice, we can then fix a *sequence of spectral cutoffs*

$$\{K(n)\}_{n\in\mathbb{N}} \subseteq \mathbb{N}$$

by prescribing that $K(n)$ is the last index of the multiplicity block of the continuum eigenvalue $\lambda_{\widetilde{K}(n)}$ for every $n \in \mathbb{N}$. Defining the cutoffs in this way allows for consistent definition of low-pass signals based on the eigenvalue magnitude, which in turn allows for changes of bases in the respective eigenspaces if needed.

2.2 Shallow graph convolutional neural networks

This subsection is dedicated to formulating convolution on graphs and manifolds, specifying the assumptions on the activation function, and defining shallow GCNNs as functionals on the space of graph signals. We will review how such functionals can either be nonlinearly parametrized by the parameters themselves or linearly by a measure in the form of a parameter distribution. The latter approach not only renders the parametrization linear, but is also flexible in the width of the neural network. This point of view is made compatible with finite realizable networks through the use of convex representer theorems guaranteeing the existence of minimizers supported on finitely many parameters, when considering a regularized empirical risk minimization cost involving finitely many training pairs. In particular, these apply to the discrete and continuum training problems we consider in Section 5, see Proposition 5.1.

To define a continuum problem on manifolds that reflects the limit behavior of GCNNs, we need to define an appropriate set of continuum signals. A natural choice in light of the spectral definition of the convolution, but which also works well with spatial discretizations, is to define a *manifold signal* to be a function $u \in L^2(\mathcal{M}, \mu)$. A *graph signal* assigns every node of the graph of size n a value in $\mathbb{R}$, so can be represented as a vector in $\mathbb{R}^n$, equipped with the standard Euclidean topology, which is isometrically isomorphic to $L^2(\mathcal{M}_n, \mu_n)$. Therefore we consider a *graph signal* to be a function $u \in L^2(\mathcal{M}_n, \mu_n)$. For readability, we will occasionally write L^2 instead of $L^2(\mathcal{M}, \mu)$, especially in subscripts of norms and inner products.

Since in general there is no group structure on graphs or manifolds, motivated by the Convolution Theorem for Euclidean domains, the convolution operations on graphs and manifolds are defined as multiplication in the respective spectral domain, i.e. given as the bilinear operations

$$\ast_n : L^2(\mathcal{M}_n, \mu_n) \times L^2(\mathcal{M}_n, \mu_n) \to L^2(\mathcal{M}_n, \mu_n), \qquad (u, v) \mapsto u \ast_n v := \sum_{k=1}^{n} \langle u, \phi_k^{[n]} \rangle_{\ell^2} \langle v, \phi_k^{[n]} \rangle_{\ell^2} \phi_k^{[n]},$$

$$\ast : L^2(\mathcal{M}, \mu) \times L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}, \mu), \qquad (u, v) \mapsto u \ast v := \sum_{k=1}^{\infty} \langle u, \phi_k \rangle_{L^2} \langle v, \phi_k \rangle_{L^2} \phi_k.$$

Let the activation function $\sigma : \mathbb{R} \to \mathbb{R}$ be a nonlinear function. We understand its application $\sigma(v)$ to a $v \in L^2(\mathcal{M}, \mu)$ as a map $\sigma(v) : \mathcal{M} \to \mathbb{R}$, $x \mapsto \sigma(v(x))$. Accordingly its application to a ℓ^2-sequence $\{v_i\}_i \in \ell^2$ is component-wise, i.e. $\sigma(\{v_i\}_i) = \{\sigma(v_i)\}_i$. We further demand from σ to have Lipschitz property with Lipschitz constant L_σ, so we have that

$$\|\sigma(v) - \sigma(\tilde{v})\|_{L^2(\mathcal{M}, \mu)} \le L_\sigma \|v - \tilde{v}\|_{L^2(\mathcal{M}, \mu)} \text{ for all } v, \tilde{v} \in L^2(\mathcal{M}, \mu). \tag{2.9}$$

Note, that this also yields a constant $C_\sigma > 0$, such that for all $u \in L^2(\mathcal{M}, \mu)$ it is

$$\|\sigma(u)\|_{L^2} \le \|\sigma(u) - \sigma(0)\|_{L^2} + \|\sigma(0)\|_{L^2} \le L_\sigma \|u\|_{L^2} + \|\sigma(0)\|_{L^2} \le C_\sigma (\|u\|_{L^2} + 1). \tag{2.10}$$

Then a trained graph convolutional neural network is a functional

$$f_\theta^{[n]} : L^2(\mathcal{M}_n, \mu_n) \to \mathbb{R}, \quad f_\theta^{[n]}(u) = \sum_{j=1}^{M} \langle a_j, \sigma(b_j \ast_n u + c_j) \rangle,$$

where $\theta = \{(a_j, b_j, c_j)\}_{j=1}^{M} \in \Theta_n^M \subset (\mathbb{R}^n)^M$ are minimizers of a regularized Empirical Risk Minimization objective

$$\arg\min_\theta \left[\frac{1}{l} \sum_{k=1}^{l} \ell \left(f_\theta^{[n]}(u_k^{[n]}), y_k \right) + R(\theta) \right].$$

Clearly, those $f_\theta^{[n]}$ are nonlinearly parametrized by their parameters θ, however parametrizing by a measure over the parameter space instead leads to defining *graph convolutional neural network* of potentially infinite width as functionals evaluating graph signals as follows

$$f_{\rho_n}^{[n]} : L^2(\mathcal{M}_n, \mu_n) \to \mathbb{R}, \quad f_{\rho_n}^{[n]}(u) = \int_{\Theta_n} \psi_n(u, \theta) d\rho_n(\theta) = \int_{\Theta_n} a^T \sigma(b \ast_n u + c) d\rho_n(a, b, c), \tag{2.11}$$

and accordingly *manifold convolutional neural networks* as functionals evaluating manifold signals as follows

$$f_\rho : L^2(\mathcal{M}, \mu) \to \mathbb{R}, \quad f_\rho(u) = \int_\Theta \psi(u, \theta) d\rho(\theta) = \int_\Theta \langle a, \sigma(b \ast u + c) \rangle_{L^2} d\rho(a, b, c). \tag{2.12}$$

Here $\Theta_n \subset \left(L^2(\mathcal{M}_n, \mu_n) \right)^3$ and $\Theta \subset \left(L^2(\mathcal{M}, \mu) \right)^3$ denote the respective parameter spaces and ρ_n and ρ are measures on Θ_n and Θ, respectively. We call the maps $\psi_n : L^2(\mathcal{M}_n, \mu_n) \times \Theta_n \to \mathbb{R}$, and $\psi : L^2(\mathcal{M}, \mu) \times \Theta \to \mathbb{R}$ arising here, given as

$$\psi_n(u, \theta) = \psi_n(u, (a, b, c)) := \langle a, \sigma(b \ast_n u + c) \rangle_{L^2(\mathcal{M}_n, \mu_n)}, \tag{2.13}$$

and

$$\psi(u, \theta) = \psi(u, (a, b, c)) := \langle a, \sigma(b \ast u + c) \rangle_{L^2(\mathcal{M}, \mu)} \tag{2.14}$$

in the following *neural response maps*.

11

2.3 Parameter spaces

Since the activation function is Lipschitz, the neural response maps $\psi(u,\cdot)$ are bounded on bounded parameter spaces for every manifold signal u. Hence, if the measure ρ is a finite Radon measure supported on a bounded parameter space Θ, then $f_\rho(u)$ is finite for every signal u. In particular, the notion of a manifold convolutional neural network linearly parametrized by ρ is well-defined.

In order to study convergence of such shallow GCNNs towards their manifold counterpart, it is useful to employ tools from functional analysis. To this end, in light of the Riesz-Markov duality, we want to interpret $f_\rho(u)$ as a dual pairing between the neural response map $\psi(u,\cdot)$ and the parametrizing measure ρ.

Such usage of Riesz-Markov duality is tied to compactness in the parameter space Θ, either by using C_0 functions in the closure of the compactly supported ones, or by requiring compactness of Θ outright. However, and in the infinite dimensional Hilbert space $\left(L^2(\mathcal{M},\mu)\right)^3$ equipped with the norm topology, compactness is not implied by boundedness.

One option would be to try and recover compactness using Banach-Alaoglu by considering its unit ball equipped with the weak topology. This would then require the neural response maps $\psi(u,\cdot)$ for fixed signals $u \in L^2(\mathcal{M},\mu)$ to be sequentially continuous with respect to the weak topology of $\left(L^2(\mathcal{M},\mu)\right)^3$. This is not compatible with the neural network setting, since given a weakly converging sequence (a_k, b_k, c_k), in the ensuing sequence of responses $\langle a_k, \sigma(b_k * u + c_k)\rangle_{L^2}$ the argument $b_k * u + c_k$ would only converge weakly, and the pointwise nonlinearity σ is never weakly continuous. Moreover, even if such an application of σ would be sequentially weakly continuous (which would force it to be linear), we would still encounter an L^2-inner product of two weakly converging sequences.

Note however, that the convolution term behaves well with respect to weak convergence. If $b_k \xrightarrow{k\to\infty} b$ in $L^2(\mathcal{M},\mu)$ implies for all $u \in L^2(\mathcal{M},\mu)$

$$\|b_k * u - b * u\|_{L^2} = \left\| \sum_{j=1}^\infty \langle b_k - b, \phi_j\rangle_{L^2} \langle u, \phi_j\rangle_{L^2} \phi_j \right\|_{L^2} \xrightarrow{j\to\infty} 0.$$

Therefore, we can resolve this and recover continuity of the neural response maps by restricting the continuum parameter space to be compact in the norm topology just in the last argument, which we can ensure this by assuming parameters to have H^α-regularity for $\alpha \in (0,1]$ in this last argument. As we will see later, for obtaining also uniform convergence of the neural response maps, we need control of differences in L^2-norm on the first parameter, so we impose the same H^α-regularity on it as well.

Following Proposition 3.2 in [10] we can write the H^α semi-norm also in the following spectral manner:

$$[v]^2_{H^\alpha(\mathcal{M},\mu)} = \sum_{k=1}^\infty \lambda_k^\alpha \langle v, \phi_k\rangle^2_{L^2(\mathcal{M},\mu)},$$

and $H^\alpha(\mathcal{M},\mu)$ forms a Hilbert space with the inner product $\langle\cdot,\cdot\rangle_{H^\alpha(\mathcal{M},\mu)}$ given as

$$\langle u, v\rangle_{H^\alpha(\mathcal{M},\mu)} := \sum_{k=1}^\infty \left(1 + \sqrt{\lambda_k}\right)^{2\alpha} \langle u, \phi_k\rangle_{L^2}\langle v, \phi_k\rangle_{L^2}. \tag{2.15}$$

Of course there is not naturally a notion of H^α regularity on the finite dimensional parameter spaces, since all norms are equivalent. Still, we want to impose a decay of the Fourier coefficients with increasing frequency, analogously to the continuum case and in a way which is convenient to

pass to the limit. So we employ this spectral definition of the semi-norm to define the following norm on $\mathbb{R}^n$:

$$\|v\|_{\alpha,n} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}, \|v\|_{\alpha,n}^2 = \sum_{k=1}^{n} \left(1 + \sqrt{\lambda_k^{[n]}}\right)^\alpha \langle v, \phi_k^{[n]} \rangle^2, \tag{2.16}$$

induced by the inner product $\langle \cdot, \cdot \rangle_{H^\alpha(\mathcal{M}_n, \mu_n)}$ on $L^2(\mathcal{M}_n, \mu_n)$ by

$$\langle u, v \rangle_{H^\alpha(\mathcal{M}_n, \mu_n)} := \sum_{k=1}^{n} \left(1 + \sqrt{\lambda_k^{[n]}}\right)^{2\alpha} \langle u, \phi_k^{[n]} \rangle_{L^2} \langle v, \phi_k^{[n]} \rangle_{L^2}.$$

Furthermore, motivated by Theorem 3.1, we impose the spectral cutoff-condition $\langle v, \phi_k^{[n]} \rangle = 0$ for all $k > N(\lambda_{K(n)})$, i.e. we restrict ourselves to the subspaces spanned by any first $N(\lambda_{K(n)})$ eigenvectors $\{\varphi_k^{[n]}\}_{k=1}^{N(\lambda_{K(n)})}$, denoted as

$$V_n := \text{Span}\left\{\varphi_1^{[n]}, \dots \varphi_{N(\lambda_{K(n)})}^{[n]}\right\}.$$

Then, we define

$$\Theta := B_{H^\alpha} \times B_{L^2} \times B_{H^\alpha}, \tag{2.17}$$

and

$$\Theta_n := \left(B_{\|\cdot\|_{\alpha,n}} \cap V_n\right) \times \left(B_{\|\cdot\|_{0,n}} \cap V_n\right) \times \left(B_{\|\cdot\|_{\alpha,n}} \cap V_n\right), \tag{2.18}$$

where B denotes the unit ball in the indicated space with the given norm, and $\|\cdot\|_{0,n}$ just denotes the Euclidean norm.

Lemma 2.5. *For all $u \in L^2(\mathcal{M}, \mu)$ it holds that $\psi(u, \cdot) \in C(\Theta, w)$.*

Proof. Let $\{\theta\}_k \subset \Theta$ be a sequence, that converges weakly in Θ to some $\theta \in \Theta$, i.e. if we write $\theta_k = (a_k, b_k, c_k)$ for $k \in \mathbb{N}$ and $\theta = (a, b, c)$, then $a_k, c_k \xrightarrow{k \to \infty} a, c$ in $H^\alpha(\mathcal{M}, \mu)$ and $b_k \xrightarrow{k \to \infty} b$ in $L^2(\mathcal{M}, \mu)$. Then the Rellich-Kondrachov for fractional Sobolev spaces [36, Lem. 10] yields strong L^2-convergence up to a subsequence in the first and last parameter $a_{k_j}, c_{k_j} \xrightarrow{j \to \infty} a, c$ in $L^2(\mathcal{M}, \mu)$.

Let $u \in L^2(\mathcal{M}, \mu)$ arbitrary but fixed. Then the sequence $\{\psi(u, \theta_k)\}_k \subset \mathbb{R}$ is bounded, namely there is a $R \geq 0$, such that

$$|\psi(u, \theta_k)| = |\langle a_k, \sigma(b_k * u + c_k)\rangle_{L^2}| \leq \|a_k\|_{L^2} C_\sigma(\|b_k * u\|_{L^2} + \|c_k\|_{L^2}) \leq R,$$

where the last inequality follows from $\{\theta_k\}_k = \{(a_k, b_k, c_k)\}_k$ being bounded in L^2 due to the definition of Θ as the unit ball of the product space (also just because of the weak convergence in L^2)

Furthermore $\psi(u, \cdot) \in C((L^2(\mathcal{M}, \mu))^3)$, i.e. it is continuous w.r.t. the strong topology of L^2, since for every sequence $\{\tilde{\theta}_k\}_k = \{(\tilde{a}_k, \tilde{b}_k, \tilde{c}_k)\}_k \subset L^2(\mathcal{M}, \mu)^3$ converging w.r.t. strong topology on L^2 to some $\tilde{\theta} = (\tilde{a}, \tilde{b}, \tilde{c}) \in L^2(\mathcal{M}, \mu)^3$, we have

$$|\psi(u, \tilde{\theta}_k) - \psi(u, \tilde{\theta})| = |\langle \tilde{a}_k, \sigma(\tilde{b}_k * u + \tilde{c}_k)\rangle_{L^2} - \langle \tilde{a}, \sigma(\tilde{b} * u + \tilde{c})\rangle_{L^2}|$$

$$\leq \|\tilde{a}_k - \tilde{a}\|_{L^2} C_\sigma(\|\tilde{b}_k\|_{L^2} \|u\|_{L^2} + \|\tilde{c}_k\|_{L^2})$$

$$+ \|\tilde{a}\|_{L^2} L_\sigma(\|(\tilde{b} - \tilde{b}_k) * u\|_{L^2} + \|\tilde{c}_k - \tilde{c}\|_{L^2}) \xrightarrow{k \to \infty} 0.$$

Now from the boundedness of $\{\psi(u, \theta_k)\}_k \subset \mathbb{R}$ and Rellich-Kondrachov we can conclude that $\psi(u, \cdot) \in C((\Theta, w))$, i.e. $|\psi(u, \theta_k) - \psi(u, \theta)| \xrightarrow{k \to \infty} 0$: Assume $|\psi(u, \theta_k) - \psi(u, \theta)| \nrightarrow 0$ as $k \to \infty$. Thus there is an $\varepsilon > 0$ and a subsequence $\{\theta_{k_l}\}_l \subset \{\theta_k\}_k$ s.t. $|\psi(u, \theta_{k_l}) - \psi(u, \theta)| \geq \varepsilon$ for all

$l \in \mathbb{N}$. Now $\{a_{k_l}\}_l, \{c_{k_l}\}_l \subset B(H^\alpha)$, so by Rellich-Kondrachov there is another subsequence $\{l_o\}_o \subset \mathbb{N}$, s.t. $a_{k_{l_o}}, c_{k_{l_o}} \xrightarrow{o \to \infty} a, c$ strongly in L^2. But then from the continuity of $\psi(u, \cdot)$ w.r.t. the strong topology on L^2 we know that there is a $O \in \mathbb{N}$ s.t. for all $o \geq O$ it holds that $|\psi(u, \theta_{k_{l_o}}) - \psi(u, \theta)| < \varepsilon$, which contradicts the assumption and proves the claim. $\qquad\square$

Now, because of the compactness $C_0((\Theta, w)) = C((\Theta, w))$, and therefore due to Riesz-Markov the dual of the space of weakly continuous functions on Θ can be identified with the space of (finite) Radon Measures on Θ, i.e.

$$\big(C((\Theta, w))\big)^* \cong \mathcal{M}(\Theta),$$

while in the finite dimensional case the weak and all norm topologies coincide, so the dual of the space of continuous functions on Θ_n can be identified with the space of (finite) Radon Measures on Θ_n, i.e.

$$\big(C(\Theta_n)\big)^* = \big(C((\Theta_n, w))\big)^* \cong \mathcal{M}(\Theta_n).$$

Therefore we can now understand neural networks $f_\rho : L^2(\mathcal{M}, \mu) \to \mathbb{R}$, $f_{\rho_n} : L^2(\mathcal{M}_n, \mu_n) \to \mathbb{R}$ as dual pairings, i.e.

$$f_\rho(u) = \langle \rho, \psi(u, \cdot) \rangle_{\mathcal{M}(\Theta), C((\Theta, w))}, \quad f_{\rho_n}(v) = \langle \rho_n, \psi_n(v, \cdot) \rangle_{\mathcal{M}(\Theta_n), C(\Theta_n)}.$$

This duality enables us to consider weak limits of in terms of neuron evaluation functionals or weak-* limits in terms of parameter distributions.

2.4 Consistency of graph signals by comparison through transport maps

The graph and continuum GCNNs defined in the previous subsections are defined on signals on different domains, so to formulate convergence results relating them we need a notion of consistency of their inputs.

We say that a collection of graph signals $u_n \in L^2(\mathcal{M}_n, \mu_n)$, $n \in \mathbb{N}$ is *consistent across resolutions* if they are a discretization of an underlying manifold signal, namely if there exists a $u \in L^2(\mathcal{M}, \mu)$ s.t. $d_{TL^2}((\mu, u), (\mu_n, u_n)) \to 0$. In our setting, where $\mu_n \overset{*}{\rightharpoonup} \mu$ as $n \to \infty$ and there exist stagnating optimal transportation maps $T_n : \mathcal{M} \to \mathcal{M}_n$ with $(T_n)_{\#}\mu = \mu_n$, TL^2 convergence only requires of the graph signals u_n that they are such that there is a $u \in L^2(\mathcal{M}, \mu)$, such that $\|u - u_n \circ T_n\|_{L^2} \xrightarrow{n \to \infty} 0$ [23, pp. 17,18].

We will see later that graph signals arising from a spatial discretization by averaging over transport cells or a spectral discretization of a manifold signal can be examples of such consistency across resolutions, see Remark 4.6.

Define the contractive *spatial discretization maps* $\mathcal{P}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$, $n \in \mathbb{N}$ which discretize by averaging over the partition [8, p. 689] and [21, p. 840]

$$(\mathcal{P}_n v)(x_i) := n \cdot \int_{U_i^{[n]}} v(x) d\mu(x) \quad \text{for } v \in L^2(\mathcal{M}, \mu), \tag{2.19}$$

where $U_i^{[n]} := T_n^{-1}(\{x_i\})$, with $\mu(U_i^{[n]}) = \frac{1}{n}$, $i \in \{1, \dots n\}$ and the adjoint extension maps $\mathcal{P}_n^* : L^2(\mathcal{M}_n, \mu_n) \to L^2(\mathcal{M}, \mu)$, $n \in \mathbb{N}$, that extend discrete functions to piecewise constant step functions over the tesselation

$$(\mathcal{P}_n^* v)(x) := v(T_n(x)) = \sum_{i=1}^n v(x_i) \mathbb{1}_{U_i^{[n]}}(x) \quad \text{for } v \in L^2(\mathcal{M}_n, \mu_n). \tag{2.20}$$

14

Remark 2.6. From the definition of $\mathcal{P}_n^*$ we see that the notions of TL^2-convergence and convergence under spatial extension $\mathcal{P}_n^*$ for any collection of graph signals $u_n \in L^2(\mathcal{M}_n, \mu_n)$ towards a manifold signal $u \in L^2(\mathcal{M}, \mu)$ coincide, i.e.

$$d_{TL^2}((\mu, u), (\mu_n, u_n)) \xrightarrow{n \to \infty} 0 \qquad \Longleftrightarrow \qquad \|u - \mathcal{P}_n^* u_n\|_{L^2} \xrightarrow{n \to \infty} 0. \tag{2.21}$$

For completeness, we also briefly introduce the interpolation operators used in the spectral approximation results of [21] that we rely on later. For $n \in \mathbb{N}$, let $I_n : L^2(\mathcal{M}_n, \mu_n) \to \text{Lip}(\mathcal{M})$ be the *interpolation operator* from [21, p. 840], defined as

$$I_n = \Lambda_{h-2\varepsilon} \mathcal{P}_n^*, \tag{2.22}$$

where $\Lambda_r : L^2(\mathcal{M}, \mu) \to \text{Lip}(\mathcal{M})$, $r > 0$ are smoothing operators, constructed by pointwisely normalizing the integral kernel operators $\Lambda_r^0 : L^2(\mathcal{M}, \text{Vol}) \to \text{Lip}(\mathcal{M}))$, that convolve a $L^2(\mathcal{M}, \text{Vol})$-function with a radial kernel $k_r(x, y)$ in the geodesic distance with compact support on the geodesic ball $B_{\mathcal{M}}(x, r)$, see [21, p. 862]. These interpolation operators find application in comparing eigenfunctions of the graph Laplacian and the Laplace-Beltrami operator.

3   Spectral Approximations

In this section we collect the error bounds for approximation of the eigenvalues and eigenspaces required for the convergence results in Section 4. Only the low-frequency part of the spectrum of the graph Laplacian approximates that of the Laplace-Beltrami well, and we will see that the spectral cutoff sequence $\{K(n)\}_n$ introduced in Section 2.1, constructed from a sequence $\{\widetilde{K}(n)\}_n$ satisfying (2.8) ensures that the assumptions from [21, Thm. 4,5, and Cor. 1] are satisfied. In particular, $K(n)$ is indeed a reasonable choice for an upper bound of the informative spectral window, and the error bounds stated here are direct consequences of the results from [21].

The situation for the approximation of eigenspaces is more subtle, due to possible high multiplicities, and the fact that multiplicity blocks of the graph Laplacian and the Laplace–Beltrami operator need not coincide. Nevertheless, for a fixed resolution and a fixed choice of orthonormal graph eigenfunctions associated with a given Laplace-Beltrami multiplicity block, one can select an orthonormal basis of the corresponding continuum eigenspace such that the discrete eigenfunctions approximate this continuum-basis well, which is a consequence of [8, Thm. 4] and [21, Thm. 5].

This approximation error depends on the spectral gap. In this work, we are interested in the discrete-to-continuum convergence, where there are no bandwidth restrictions on the manifold signal space, hence the spectral window must ultimately diverge. Thus, achieving uniform control of eigenspace approximation errors requires assumptions on the asymptotic behavior of spectral gaps of the Laplace–Beltrami operator. We therefore impose the polynomial gap condition stated in Assumption 2.2. For more discussion on this, see also Section 6.1.

Theorem 3.1 (Error bounds for the relative eigenvalue and eigenspace approximation)**.** *Given a sequence of neighborhood radii $\{h_n\}_n$ and a sequence of spectral cutoffs $\{\widetilde{K}(n)\}_n$ subject to (2.3) and (2.8) and the corresponding sequence $\{K(n)\}_n \in \mathbb{N}$ as defined above, for all $n \in \mathbb{N}$ it holds that for every $k \leq K(n)$ that*

$$|\lambda_k^{[n]} - \lambda_k| \leq C\lambda_k \delta(\varepsilon_n, h_n, \lambda_k). \tag{3.1}$$

Furthermore for every fixed $n \in \mathbb{N}$ and a collection $\{\varphi_k^{[n]}\}_{k=1}^n$ of normalized eigenfunctions of Δ_n corresponding to the eigenvalues $\lambda_1^{[n]}, \ldots \lambda_n^{[n]}$, there exist normalized eigenfunctions $\{\varphi_k\}_{k=1}^n$ of

the eigenvalues $\lambda_1, \dots \lambda_n$ *of the Laplace-Beltrami operator* Δ *such that*

$$\left\| I_n \varphi_k^{[n]} - \varphi_k \right\|_{L^2} \leq \frac{C}{\gamma_{k,\mu}} \delta(\varepsilon_n, h_n, \lambda_k) \tag{3.2}$$

where $I_n : L^2(\mathcal{M}_n, \mu_n) \to Lip(\mathcal{M})$, $n \in \mathbb{N}$ *are the interpolation operators from* [21][p. 840], *and*

$$\delta(\varepsilon, h, \lambda) := \frac{\varepsilon}{h} + \left(1 + \sqrt{\lambda}\right) h + \left(K + \frac{1}{R^2}\right) h^2. \tag{3.3}$$

Proof. The stated bounds will be direct consequences of Theorem 4 (or in particular Corollary 1) and Theorem 5 in [21, pp. 837-840], once we ensure that their assumptions are satisfied. Weyl's law states that the number of eigenvalues (including multiplicity) smaller than a $\lambda > 0$

$$N(\lambda) := |\{k \in \mathbb{N} \mid \lambda_k \leq \lambda, \lambda_k \text{ eigenvalue of } \Delta\}| \tag{3.4}$$

asymptotically behave like

$$\lim_{\lambda \to \infty} N(\lambda) \lambda^{-m/2} = \frac{\omega_m}{(2\pi)^m} \text{Vol}_{\mathcal{M}}(\mathcal{M}) =: C_{\mathcal{M}},$$

where ω_m denotes the volume of the m-ball $B(0) \subset \mathbb{R}^m$. Specifically, we will use the more refined asymptotic law [27]

$$N(\lambda) = \frac{\omega_m}{(2\pi)^m} \text{Vol}_{\mathcal{M}}(\mathcal{M}) \lambda^{m/2} + O(\lambda^{\frac{m-1}{2}}). \tag{3.5}$$

Let $\{k_j\}_j \subset \mathbb{N}$ enumerate the first index in each multiplicity block, i.e. $\lambda_{k_j} \neq \lambda_{k_j - 1}$, and denote by $m(\lambda_k)$ the multiplicity of λ_k. Then

$$N(\lambda_k) = \sum_{j \in \mathbb{N}, k_j \leq k} m(\lambda_{k_j}) \geq k.$$

Now we have for all $\delta > 0$ and $j \in \mathbb{N}$ by (3.5) and the mean-value theorem that

$$m(\lambda_{k_j}) \leq N(\lambda_{k_j} + \delta) - N(\lambda_{k_j} - \delta)$$

$$= C_{\mathcal{M}} \left((\lambda_{k_j} + \delta)^{m/2} - (\lambda_{k_j} - \delta)^{m/2} \right) + O\left((\lambda_{k_j} + \delta)^{\frac{m-1}{2}} \right) - O\left((\lambda_{k_j} - \delta)^{\frac{m-1}{2}} \right)$$

$$\leq C_{\mathcal{M}} m \delta (\lambda_{k_j} + \delta)^{\frac{m-2}{2}} + O\left((\lambda_{k_j} + \delta)^{\frac{m-1}{2}} \right) = O\left((\lambda_{k_j} + \delta)^{\frac{m-1}{2}} \right).$$

Since $\delta > 0$ was arbitrary we may conclude that

$$m(\lambda_k) = O\left(\lambda_k^{\frac{m-1}{2}} \right) \qquad \text{for all } k \in \mathbb{N}. \tag{3.6}$$

Therefore $\lim_{k \to \infty} m(\lambda_k) \lambda_k^{-m/2} = 0$, since then

$$m(\lambda_k) \lambda_k^{-m/2} = O\left(\lambda_k^{\frac{m-1}{2}} \right) \lambda_k^{-m/2} = O\left(\lambda_k^{-1/2} \right),$$

which yields

$$\lim_{k \to \infty} k \lambda_k^{-m/2} \leq \lim_{k \to \infty} N(\lambda_k) \lambda_k^{-m/2} \leq \lim_{k \to \infty} (k + m(\lambda_k)) \lambda_k^{-m/2} = \lim_{k \to \infty} k \lambda_k^{-m/2},$$

and hence

$$C_{\mathcal{M}} = \lim_{k \to \infty} N(\lambda_k) \lambda_k^{-m/2} = \lim_{k \to \infty} k \lambda_k^{-m/2}. \tag{3.7}$$

If we denote by $\bar{k}$ the last index of the multiplicity block of λ_k, then we see that on the one hand from (3.7)

$$\lim_{k \to \infty} k \lambda_{\bar{k}}^{-m/2} \leq \lim_{k \to \infty} \bar{k} \lambda_{\bar{k}}^{-m/2} = C_{\mathcal{M}},$$

16

and on the other hand from using additionally (3.6)

$$\lim_{k\to\infty} k\lambda_{\bar{k}}^{-m/2} \geq \lim_{k\to\infty}(N(\lambda_k) - m(\lambda_{\bar{k}}))\lambda_{\bar{k}}^{-m/2} = \lim_{k\to\infty} N(\lambda_{\bar{k}})\lambda_{\bar{k}}^{-m/2} - \lim_{k\to\infty} m(\lambda_{\bar{k}})\lambda_{\bar{k}}^{-m/2} = C_{\mathcal{M}} - 0 = C_{\mathcal{M}},$$

hence it holds that

$$\lim_{k\to\infty} k\lambda_{\bar{k}}^{-m/2} = C_{\mathcal{M}}. \tag{3.8}$$

So with this and since by construction $K(n) = N(\lambda_{\tilde{K}(n)})$, as well as $(m+1)/2 \geq 1$ for all $m \geq 1$ it holds that $\tilde{K}(n)^{\frac{1}{m}} \leq \tilde{K}(n)^{\frac{1}{m}\frac{m+1}{2}}$ and thus condition (2.8) yields that for every $n \in \mathbb{N}$

$$0 \leq \lim_{n\to\infty} h_n\sqrt{\lambda_{K(n)}} = \lim_{n\to\infty} h_n\tilde{K}(n)^{1/m}\tilde{K}(n)^{-1/m}\sqrt{\lambda_{K(n)}} = \lim_{n\to\infty} h_n\tilde{K}(n)^{1/m}\lim_{k\to\infty} k^{-1/m}\sqrt{\lambda_{\bar{k}}}$$
$$\leq \lim_{n\to\infty} h_n\tilde{K}(n)^{1/m}\lim_{k\to\infty} k^{-1/m}\sqrt{\lambda_{\bar{k}}} = 0 \cdot C_{\mathcal{M}}^{-1/m} = 0, \tag{3.9}$$

so $\{(\sqrt{\lambda_{K(n)}} + 1)h_n\}_n$ is a bounded sequence converging to 0, and therefore there is a constant $c_{\mathcal{M}} > 0$ such that for all $n \in \mathbb{N}$ and $k \leq K(n)$

$$\left(\sqrt{\lambda_k} + 1\right)h_n \leq \left(\sqrt{\lambda_{K(n)}} + 1\right)h_n \leq c_{\mathcal{M}}.$$

This is precisely the assumption required for the results of [21] mentioned above. □

Corollary 3.2. *For the error bounds from Theorem 3.1, it holds that*

$$\lim_{n\to\infty} \delta(\varepsilon_n, h_n, \lambda_{K(n)}) = 0. \tag{3.10}$$

or equivalently for every $k \in \mathbb{N}$ and N_k such that $k \leq K(n)$ for all $n \geq N_k$

$$\lim_{N_k \leq n \to \infty} \delta(\varepsilon_n, h_n, \lambda_k) = 0.$$

Proof. This follows immediately from the observation (3.9) and the assumption (2.3). □

Lemma 3.3. *Denote $\Gamma_{K,\mu} := \min_{k \leq K} \gamma_{k,\mu}$, where $\gamma_{k,\mu}$ is the spectral gap defined in (2.7). Then it holds that*

$$\lim_{n\to\infty} \frac{1}{\Gamma_{K(n),\mu}}\delta(\varepsilon_n, h_n, \lambda_{K(n)})K(n)^{\frac{m-1}{2m}} = 0. \tag{3.11}$$

Proof. First, note that by definition $\Gamma_{K(n),\mu} = \Gamma_{\tilde{K}(n),\mu}$, since $\lambda_{\tilde{K}(n)} = \lambda_{\tilde{K}(n)+j}$ for all $j \in \{0, \ldots K(n) - \tilde{K}(n)\}$. If $\beta^* = 0$, then $\{\gamma_{k,\mu}^{-1}\}_k$ is a bounded sequence, and therefore so is $\{\Gamma_{K(n),\mu}^{-1}\}_n$. So the claim follows from (3.10). Hence, we can assume $\beta^* > 0$. First notice that there is a $M \geq 0$ and a $K \in \mathbb{N}$, such that for all $k \geq K$ we have $\gamma_{k,\mu}^{-1} \leq Mk^{\beta^*}$, so if we denote

$$L(n) = \arg\min_{k \leq \tilde{K}(n)} \gamma_{k,\mu} = \arg\min_{k \leq K(n)} \gamma_{k,\mu}$$

for $n \in \mathbb{N}$, then we have for all $n \in \mathbb{N}$ with $L(n) \geq K$ that

$$\Gamma_{K(n),\mu}^{-1} = \gamma_{L(n),\mu}^{-1} \leq ML(n)^{\beta^*} \leq M\tilde{K}(n)^{\beta^*}.$$

Since $\beta^* > 0$ it is $\liminf_{k\to\infty} \gamma_{k,\mu} = 0$, so indeed there can only be finitely many such $n \in \mathbb{N}$ where $L(n) < K$. Therefore we get

$$0 \leq \lim_{n\to\infty} \frac{1}{\Gamma_{K(n),\mu}}\tilde{K}(n)^{-\beta^*}\tilde{K}(n)^{\beta^*}\delta(\varepsilon_n, h_n, \lambda_{K(n)})\tilde{K}(n)^{\frac{m-1}{2m}}$$

17

$$\leq M \lim_{n\to\infty} \tilde{K}(n)^{\beta^* + \frac{m-1}{2m}} \left(\frac{\varepsilon_n}{h_n} + \left(1 + \sqrt{\lambda_{K(n)}}\right) h_n + \left(K + \frac{1}{R^2}\right) h_n^2 \right).$$

Now the conditions (2.8) yield

$$\lim_{n\to\infty} h_n^2 \tilde{K}(n)^{\beta^* + \frac{m-1}{2m}} = \lim_{n\to\infty} h_n \tilde{K}(n)^{\beta^* + \frac{m+1}{2m}} \tilde{K}(n)^{-1/m} h_n$$

$$= \lim_{n\to\infty} h_n \tilde{K}(n)^{\beta^* + \frac{m-1}{2m}} \lim_{n\to\infty} h_n \lim_{n\to\infty} \tilde{K}(n)^{-1/m} = 0,$$

as well as

$$\lim_{n\to\infty} h_n \sqrt{\lambda_{K(n)}} \tilde{K}(n)^{\beta^* + \frac{m-1}{2m}} = \lim_{n\to\infty} h_n \tilde{K}(n)^{\beta^* + \frac{m-1}{2m}} \tilde{K}(n)^{1/m} \tilde{K}(n)^{-1/m} \sqrt{\lambda_{K(n)}}$$

$$= \lim_{n\to\infty} h_n \tilde{K}(n)^{\beta^* + \frac{m+1}{2m}} \lim_{n\to\infty} \tilde{K}(n)^{-1/m} \sqrt{\lambda_{K(n)}} = 0 \cdot C_{\mathcal{M}}^{-1/m} = 0,$$

and finally

$$\lim_{n\to\infty} \frac{\varepsilon_n}{h_n} \tilde{K}(n)^{\beta^* + \frac{m-1}{2m}} = 0. \qquad \square$$

Corollary 3.4. *Denote* $\Gamma_{K,\mu} := \min_{k \leq K} \gamma_{k,\mu}$, *where* $\gamma_{k,\mu}$ *is the spectral gap defined in* (2.7). *Then it holds that*

$$\lim_{n\to\infty} \frac{1}{\Gamma_{K(n),\mu}} \delta(\varepsilon_n, h_n, \lambda_{K(n)}) = 0. \tag{3.12}$$

Proof. This follows immediately from (3.11). $\qquad \square$

Lemma 3.5. *For* $k \in \mathbb{N}$ *denote by* $\underline{k}$ *and* $\bar{k} = \underline{k} + m(\lambda_k) - 1$ *the first and last index of the multiplicity block of* λ_k, *i.e.* $\lambda_{\underline{k}-1} < \lambda_{\underline{k}} = \ldots \lambda_k = \ldots = \lambda_{\bar{k}} < \lambda_{\bar{k}+1}$. *Let* $N \in \mathbb{N}$ *such that* $k \leq \bar{k} \leq K(N)$. *Fix* $n \geq N$ *and denote by* $\{\varphi_j^{[n]}\}_{j=\underline{k}}^{\bar{k}}$ *a choice of orthonormal eigenfunctions of the n-th Graph-Laplacian* Δ_n *corresponding to the eigenvalues* $\{\lambda_i^{[n]}\}_{i=\underline{k}}^{\bar{k}}$. *Then there exists an orthonormal basis* $\{\varphi_j^{(n)}\}_{j=\underline{k}}^{\bar{k}}$ *of the Laplace-Beltrami eigenspace* $E(\lambda_k)$ *such that for all* $j \in \{\underline{k}, \ldots \bar{k}\}$

$$\left\| I_n \varphi_j^{[n]} - \varphi_j^{(n)} \right\|_{L^2} \leq \frac{C}{\Gamma_{K(n),\mu}} \delta(\varepsilon_n, h_n, \lambda_{K(n)}) \tag{3.13}$$

and

$$\left\| \mathcal{P}_n^* \varphi_j^{[n]} - \varphi_j^{(n)} \right\|_{L^2} \leq C \left(h_n \sqrt{\lambda_{K(n)}} + \frac{1}{\Gamma_{K(n),\mu}} \delta(\varepsilon_n, h_n, \lambda_{K(n)}) \right) =: \delta_n^\phi \tag{3.14}$$

Proof. The existence of orthonormal eigenfunctions $\{\varphi_j^{(n)}\}_{j=\underline{k}}^{\bar{k}}$ for a choice of $\{\varphi_j^{[n]}\}_{j=\underline{k}}^{\bar{k}}$ of orthonormal eigenfunctions of the n-th Graph-Laplacian satisfying (3.13) follows from [21, Thm. 5] and [8, Thm. 4].

Then following [21, p. 879] and using Theorem 3.1, in particular (3.2), we get

$$\left\| \mathcal{P}_n^* \varphi_j^{[n]} - \varphi_j^{(n)} \right\|_{L^2} \leq \left(\left\| I_n \varphi_j^{[n]} - \mathcal{P}_n^* \varphi_j^{[n]} \right\|_{L^2} + \left\| I_n \varphi_j^{[n]} - \varphi_j^{(n)} \right\|_{L^2} \right)$$

$$\leq C h_n \sqrt{\lambda_k} + \frac{C}{\gamma_{k,\mu}} \delta(\varepsilon_n, h_n, \lambda_k)$$

$$\leq C \left(h_n \sqrt{\lambda_{K(n)}} + \frac{1}{\Gamma_{K(n),\mu}} \delta(\varepsilon_n, h_n, \lambda_{K(n)}) \right). \qquad \square$$

Note, that the choice of eigenfunctions of the Laplace-Beltrami operator in Lemma 3.5 depend on the choice of eigenfunctions of the Graph-Laplacian for a fixed resolution n.

4 Continuum-to-discrete projections

In order to show consistency of the graph networks across resolutions, we need some ways to meaningfully compare objects in the discrete and continuum spaces. A difficulty in comparing discrete neural response maps ψ_n to the continuum map ψ is that the activation function acts pointwise, while the convolution operation is defined spectrally.

In the previous Section 2.4 we already defined a spatial discretization $\mathcal{P}_n$ of L^2-functions, through averaging over optimal transportation cells. While this yields a natural approximation compatible with both the sampling and the L^2-structure, quantifying the loss of information by such a discretization is problematic when dealing with spectral operations. In contrast, the normalized eigenfunctions of the Laplace-Beltrami operator and the orthonormal eigenvectors of the graph Laplacians provide bases for the respective L^2-spaces, which allows for the definition of spectral projection maps that preserve the Hilbert structure of L^2 and its Sobolev subspaces.

This section first introduces such spectral projections, and subsequently concerns itself with the properties of both spatial and spectral projections, and their interplay. Neither projection is generally better suited for the discrete-to-continuum analysis of nonlinear convolutional neural response maps. Rather, they serve complementary roles. On the one hand, convolutions are expressed directly as multiplications in the spectral domain. And on the other hand, the spatial extension maps, being constant on each transport cell, commute with the pointwise application of the activation function, which is crucial for handling the nonlinearity.

Equipped with these projection mappings, at the end of this section we obtain pointwise convergence of the neural response maps ψ_n to ψ under appropriate discretization of the parameters and the signal in Corollary 4.18. As a consequence of the tension between their spatial and spectral aspects, and the fact that graph convolution is based on the respective graph eigenfunctions, the discrete expression has to be lifted spatially, whereas asymptotic control of convolution operations requires the introduction of spectral projections. This leads to controlling the asymptotic behavior of expressions such as

$$\mathcal{P}_n^*(\mathcal{S}_n b *_n \mathcal{R}_n u) - \mathcal{S}_n^* \mathcal{S}_n b * \mathcal{P}_n^* \mathcal{R}_n u,$$

appearing in (4.16). Here, $\mathcal{P}_n^*$ denotes the spatial extension map from (2.20), $\mathcal{S}_n$ a spectral discretization map with adjoint $\mathcal{S}_n^*$ and $\mathcal{R}_n$ generic discretization maps, that are compatible with TL^2-convergence and uniformly bounded in n. The latter are not made specific because they are reflected in conditions for the training data, which we want to restrict as little as possible. As can be seen from this, there appears a melange of discretizations and extension operators, so in Section 4.2 we will give all the convergence results required to treat such interactions.

4.1 Spectral projections and their adjoints

We first introduce the spectral projections from continuum to discrete signals used in the sequel, their adjoints which serve as extension operators

For any $\alpha \in [0,1]$ we introduce the H^α-*spectral projection maps* $\mathcal{S}_{n,\alpha} : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$, $n \in \mathbb{N}$, given as follows: For every $n \in \mathbb{N}$ fix orthonormal eigenfunctions of the n-th graph Laplacian $\{\phi_k^{[n]}\}_{k=1}^{K(n)}$ corresponding to the first $K(n)$ discrete eigenvalues. Then we can find for every $n \in \mathbb{N}$ by Lemma 3.5 continuum eigenfunctions $\{\phi_k^{(n)}\}_{k=1}^{K(n)}$ such that

$$\left\| I_n \phi_k^{[n]} - \phi_k^{(n)} \right\|_{L^2} \le \frac{C}{\Gamma_{K(n),\mu}} \delta(\varepsilon_n, h_n, \lambda_{K(n)})$$

and

$$\left\| \mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)} \right\|_{L^2} \le C h_n \sqrt{\lambda_{K(n)}} + \frac{C}{\Gamma_{K(n),\mu}} \delta(\varepsilon_n, h_n, \lambda_{K(n)}) =: \delta_n^\phi.$$

Note that we first fix a choice of graph eigenfunctions $\{\phi_k^{[n]}\}_{k=1}^{K(n)}$ for every resolution, to then be able to choose for every resolution a collection of continuum eigenfunctions $\{\phi_k^{(n)}\}_{k=1}^{K(n)}$ that aligns well with the choice for the graph of that resolution.

From this fixed choice of collections of eigenfunctions of the graph-Laplacian and the Laplace-Beltrami operator, we then define $S_{n,\alpha}$ such that for $v \in L^2(\mathcal{M}, \mu)$ and $n \in \mathbb{N}$

$$\mathcal{S}_{n,\alpha} v := \sum_{k=1}^{K(n)} \left(\frac{1 + \sqrt{\lambda_k}}{1 + \sqrt{\lambda_k^{[n]}}} \right)^{\alpha} \langle v, \phi_k^{(n)} \rangle_{L^2(\mathcal{M},\mu)} \varphi_k^{[n]}, \tag{4.1}$$

so $S_{n,\alpha}$ discretizes continuum signals by retaining only the first $K(n)$ Laplace-Beltrami modes in the resolution-dependent orthonormal basis $\{\phi_k^{(n)}\}$ induced by the chosen orthonormal graph eigenbasis, scaling the associated spectral coefficients according to the regularity parameter α and projecting back to the discrete spatial domain via said eigenvectors of the n-th graph Laplacian.

Their adjoint extension map $\mathcal{S}_{n,\alpha}^* : L^2(\mathcal{M}_n, \mu_n) \to L^2(\mathcal{M}, \mu)$, $n \in \mathbb{N}$ are given by

$$\mathcal{S}_{n,\alpha}^* w = \sum_{k=1}^{K(n)} \left(\frac{1 + \sqrt{\lambda_k^{[n]}}}{1 + \sqrt{\lambda_k}} \right)^{\alpha} \langle w, \phi_k^{[n]} \rangle_{L^2(\mathcal{M}_n, \mu_n)} \phi_k^{(n)}. \tag{4.2}$$

We write shorthand $\mathcal{S}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$, and $\mathcal{S}_n^* : L^2(\mathcal{M}_n, \mu_n) \to L^2(\mathcal{M}, \mu)$ for $\mathcal{S}_{n,0}$ and $\mathcal{S}_{n,0}^*$, $n \in \mathbb{N}$, respectively.

Denote by $\mathbb{P}(V, w)$ the $L^2(\mathcal{M}, \mu)$-orthogonal projection of $w \in L^2(\mathcal{M}, \mu)$ onto the closed subspace $V \subset L^2(\mathcal{M}, \mu)$ and by $E(\lambda)$ the eigenspace of the Laplace-Beltrami operator associated to the eigenvalue $\lambda \geq 0$. In the following, we will denote by $\{\varphi_k\}_k$ an arbitrary orthonornmal basis of $L^2(\mathcal{M}, \mu)$ consisting of eigenfunctions of the Laplace-Beltrami operator, such that for any $k \in \mathbb{N}$ the eigenfunction φ_k corresponds to the eigenvalue λ_k. Note that for any such choice and any $w \in L^2(\mathcal{M}, \mu)$ it holds that

$$\mathbb{P}\left(\bigoplus_{k=1}^{K(n)} E(\lambda_k), w \right) = \sum_{k=1}^{K(n)} \langle w, \varphi_k \rangle_{L^2(\mathcal{M},\mu)} \varphi_k = \sum_{k=1}^{K(n)} \langle w, \phi_k^{(n)} \rangle_{L^2(\mathcal{M},\mu)} \phi_k^{(n)}, \tag{4.3}$$

due to the orthogonality of distinct eigenspaces, and since the orthogonal projection onto each eigenspace is independent of the choice of orthonormal basis.

4.2   Properties of the continuum-to-discrete projections

Here we will prove properties for each of the continuum-to-discrete projections separately, but also jointly and in relation to convolutions and graph convolutions, all of which are required to show Lemma 4.11. Many of these results involve convergence statements as $n \to \infty$, and for this first set of results it might be possible to obtain convergence rates, but we do not do so because those would be lost in the arguments involving nonlinearities and Γ-convergence of later sections.

Lemma 4.1 (Properties of the Averaging Discretization and Stepwise Extension)**.** *The following properties hold for all $n \in \mathbb{N}$ for the maps $\mathcal{P}_n$ and $\mathcal{P}_n^*$:*

i) It holds that $\mathcal{P}_n$ and $\mathcal{P}_n^*$ are L^2-adjoint, that is

$$\langle \mathcal{P}_n u, v \rangle_{L^2(\mathcal{M}_n, \mu_n)} = \langle u, \mathcal{P}_n^* v \rangle_{L^2(\mathcal{M}, \mu)} \quad \text{for all } u \in L^2(\mathcal{M}, \mu), v \in L^2(\mathcal{M}_n, \mu_n).$$

ii) The extension map $\mathcal{P}_n^*$ preserves the L^2-inner product, i.e.

$$\langle u, v \rangle_{L^2(\mathcal{M}_n, \mu_n)} = \langle \mathcal{P}_n^* u, \mathcal{P}_n^* v \rangle_{L^2(\mathcal{M}, \mu)} \quad \text{for all } u, v \in L^2(\mathcal{M}_n, \mu_n).$$

iii) The extension map $\mathcal{P}_n^*$ preserves the L^2-norm, i.e.

$$\|\mathcal{P}_n^* v\|_{L^2(\mathcal{M}, \mu)} = \|v\|_{L^2(\mathcal{M}_n, \mu_n)} \quad \text{for all } v \in L^2(\mathcal{M}_n, \mu_n).$$

iv) The operator $\mathcal{P}_n^* \mathcal{P}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}, \mu)$ is linear and bounded with operator norm $\|\mathcal{P}_n^* \mathcal{P}_n\| = 1$. In particular, it holds that $\mathcal{P}_n^* \mathcal{P}_n$ is contracting in the L^2-norm, i.e.

$$\|\mathcal{P}_n^* \mathcal{P}_n v\|_{L^2(\mathcal{M}, \mu)} \leq \|v\|_{L^2(\mathcal{M}, \mu)} \quad \text{for all } v \in L^2(\mathcal{M}, \mu).$$

v) The maps $\mathcal{P}_n$, $\mathcal{P}_n^*$ are linear.
vi) It holds that $\big(\sigma(\mathcal{P}_n^* v)\big)(x) = \mathcal{P}_n^*\big(\sigma(v)\big)(x)$ for any $v \in L^2(\mathcal{M}_n, \mu_n)$ and $x \in M$.

Proof. (i), (ii) and (v) follow immediately from the definitions and imply (iii).

(iv) Linearity follows immediately from the definition of $\mathcal{P}_n$ and $\mathcal{P}_n^*$. Fix an arbitrary $v \in L^2(\mathcal{M}, \mu)$. Then

$$\|\mathcal{P}_n^* \mathcal{P}_n v\|_{L^2(\mathcal{M}, \mu)} = \int_{\mathcal{M}} |(\mathcal{P}_n^* \mathcal{P}_n v)(x)|^2 \, d\mu(x) = \sum_{i=1}^n \int_{U_i^{[n]}} |\mathcal{P}_n v(x_i)|^2 \, d\mu(x)$$

$$= \sum_{i=1}^n \int_{U_i^{[n]}} \left| \frac{1}{\mu(U_i^{[n]})} \int_{U_i^{[n]}} v(y) d\mu(y) \right|^2 d\mu(x)$$

$$= \sum_{i=1}^n \mu(U_i^{[n]}) \left| \frac{1}{\mu(U_i^{[n]})} \int_{U_i^{[n]}} v(y) d\mu(y) \right|^2$$

$$(\text{Jensen's ineq}) \leq \sum_{i=1}^n \frac{\mu(U_i^{[n]})}{\mu(U_i^{[n]})} \int_{U_i^{[n]}} |v(y)|^2 \, d\mu(y) = \int_{\mathcal{M}} |v(y)|^2 \, d\mu(y) = \|v\|_{L^2(\mathcal{M}, \mu)}^2.$$

But on the other hand, let $\{c_i\}_{i=1}^n \subset \mathbb{R}$, and define $v \in L^2(\mathcal{M}, \mu)$ as $v(x) = \sum_{i=1}^n c_i \mathbb{1}_{U_i^{[n]}}(x)$. Then it is easy to check that $\mathcal{P}_n^* \mathcal{P}_n v(x) = v(x)$ for all $x \in \mathcal{M}$, thus

$$\|(\mathcal{P}_n^* \mathcal{P}_n)v\|_{L^2(\mathcal{M}, \mu)} = \|v\|_{L^2(\mathcal{M}, \mu)},$$

yielding operator norm $\|\mathcal{P}_n^* \mathcal{P}_n\| = 1$.

(vi) Writing out the definitions yields

$$\big(\sigma(\mathcal{P}_n^* v)\big)(x) = \sigma\big((\mathcal{P}_n^* v)(x)\big) = \sigma\Big(\sum_{i=1}^n v(x_i) \mathbb{1}_{U_i^{[n]}}(x)\Big) = \sum_{i=1}^n \sigma(v(x_i)) \mathbb{1}_{U_i^{[n]}}(x) = \mathcal{P}_n^*\big(\sigma(v)\big)(x).$$

The second to last equality holds, despite the non-linearity of σ, because of the disjoint tesselation by the $U_i^{[n]}, i \in \{1, \dots n\}$. $\qquad \square$

Lemma 4.2. *For any $v \in L^2(\mathcal{M}, \mu)$ it holds that $\|\mathcal{P}_n^* \mathcal{P}_n v - v\|_{L^2(\mathcal{M}, \mu)} \xrightarrow{n \to \infty} 0$.*

Proof. Fix an arbitrary $v \in L^2(\mathcal{M}, \mu)$ and an $\varepsilon > 0$. By density of $C_c(\mathcal{M})$ in $L^2(\mathcal{M}, \mu)$, there exists $v_\varepsilon \in C_c(\mathcal{M})$ with $\|v - v_\varepsilon\|_{L^2(\mathcal{M}, \mu)} < \varepsilon$. From Lemma 4.1.(iv) it follows that

$$\|\mathcal{P}_n^* \mathcal{P}_n v_\varepsilon - \mathcal{P}_n^* \mathcal{P}_n v\|_{L^2(\mathcal{M}, \mu)} = \|\mathcal{P}_n^* \mathcal{P}_n (v_\varepsilon - v)\|_{L^2(\mathcal{M}, \mu)} \leq \|v_\varepsilon - v\|_{L^2(\mathcal{M}, \mu)} < \varepsilon.$$

Furthermore we have that

$$\|v_\varepsilon - \mathcal{P}_n^* \mathcal{P}_n v_\varepsilon\|_{L^2(\mathcal{M},\mu)}^2 = \int_{\mathcal{M}} |(v_\varepsilon - \mathcal{P}_n^* \mathcal{P}_n v_\varepsilon)(x)|^2 \, d\mu(x)$$

$$= \sum_{i=1}^n \int_{U_i^{[n]}} |v_\varepsilon(x) - \mathcal{P}_n v_\varepsilon(x_i)|^2 \, d\mu(x)$$

$$= \sum_{i=1}^n \int_{U_i^{[n]}} \left| v_\varepsilon(x) - \frac{1}{\mu(U_i^{[n]})} \int_{U_i^{[n]}} v_\varepsilon(y) d\mu(y) \right|^2 \, d\mu(x)$$

$$= \sum_{i=1}^n \int_{U_i^{[n]}} \left| \frac{1}{\mu(U_i^{[n]})} \int_{U_i^{[n]}} v_\varepsilon(x) - v_\varepsilon(y) d\mu(y) \right|^2 \, d\mu(x)$$

$$\text{(Jensen's ineq)} \leq \sum_{i=1}^n \frac{1}{\mu(U_i^{[n]})} \int_{U_i^{[n]}} \int_{U_i^{[n]}} |v_\varepsilon(x) - v_\varepsilon(y)|^2 \, d\mu(y) d\mu(x)$$

Since the support of v_ε is compact, it is actually uniformly continuous. Now, we note that

$$\varepsilon_n = d_\infty(\mu, \mu_n) = \min_{T: T_\# \mu = \mu_n} \text{esssup}_{x \in \mathcal{M}} \, d(x, T(x))$$

is attained at some map T_n. This follows from the fact that sets of the form $\{y \in \mathcal{M} : d(y,x) = t\}$ have $\mathcal{H}^m$ measure zero, which implies that any measurable map assigning Voronoi cells $U_i^{[n]}$ to their centers is optimal. Using these maps, we get

$$\text{diam}_{\mathcal{M}}(U_i^{[n]}) = \sup_{x,y \in U_i^{[n]}} d(x,y) \leq \sup_{x,y \in U_i^{[n]}} [d(x,T_n(x)) + d(y,T_n(y))] = 2 \sup_{x \in U_i^{[n]}} d(x,T_n(x))$$

$$\leq 2 \, \text{esssup}_{x \in \mathcal{M}} \, d(x, T_n(x)) = 2 \, d_\infty(\mu, \mu_n) = 2\varepsilon_n. \tag{4.4}$$

Therefore, $\text{diam}_{\mathcal{M}}(U_i^{[n]}) \to 0$ for all $i = 1, \ldots, n$ as $n \to \infty$. Thus in particular, there is a $N \in \mathbb{N}$, such that for all $n \geq N$ and for all $x, y \in U_i^{[n]}$, $i = 1, \ldots, n$ it holds that $|v_\varepsilon(x) - v_\varepsilon(y)| < \varepsilon$.

Hence for all $n \geq N$

$$\|v_\varepsilon - \mathcal{P}_n^* \mathcal{P}_n v_\varepsilon\|_{L^2(\mathcal{M},\mu)}^2 \leq \sum_{i=1}^n \frac{1}{\mu(U_i^{[n]})} \int_{U_i^{[n]}} \int_{U_i^{[n]}} |v_\varepsilon(x) - v_\varepsilon(y)|^2 \, d\mu(y) d\mu(x)$$

$$< \sum_{i=1}^n \frac{1}{\mu(U_i^{[n]})} \int_{U_i^{[n]}} \int_{U_i^{[n]}} \varepsilon^2 d\mu(y) d\mu(x)$$

$$= \sum_{i=1}^n \frac{\mu(U_i^{[n]})^2}{\mu(U_i^{[n]})} \varepsilon^2 = \sum_{i=1}^n \mu(U_i^{[n]})\varepsilon = \frac{1}{n} n\varepsilon^2 = \varepsilon^2$$

Combining the above yields

$$\|v - \mathcal{P}_n^* \mathcal{P}_n v\|_{L^2} \leq \|v - v_\varepsilon\|_{L^2} + \|v_\varepsilon - \mathcal{P}_n^* \mathcal{P}_n v_\varepsilon\|_{L^2} + \|\mathcal{P}_n^* \mathcal{P}_n v_\varepsilon - \mathcal{P}_n^* \mathcal{P}_n v\|_{L^2} < 3\varepsilon.$$

Since $\varepsilon > 0$ and $v \in L^2(\mathcal{M}, \mu)$ were arbitrary, this proves the claim. $\qquad \square$

Lemma 4.3 (Properties of the H^α Spectral Projections). *The following properties hold for all $n \in \mathbb{N}$, $\alpha \in [0,1]$ for the maps $\mathcal{S}_{n,\alpha}$ and $\mathcal{S}_{n,\alpha}^*$:*

i) The maps $\mathcal{S}_{n,\alpha}$ and $\mathcal{S}_{n,\alpha}^*$ are H^α-adjoint, that is

$$\langle \mathcal{S}_{n,\alpha} v, w \rangle_{H^\alpha(\mathcal{M}_n, \mu_n)} = \langle v, \mathcal{S}_{n,\alpha}^* w \rangle_{H^\alpha(\mathcal{M}, \mu)} \quad \text{for all } v \in L^2(\mathcal{M}, \mu), w \in L^2(\mathcal{M}_n, \mu_n).$$

ii) For all $u, v \in L^2(\mathcal{M}_n, \mu_n)$ with $\langle u, \phi_k^{[n]} \rangle_{L^2(\mathcal{M}_n, \mu_n)} = \langle v, \phi_k^{[n]} \rangle_{L^2(\mathcal{M}_n, \mu_n)} = 0$ for all $k > K(n)$ it holds that

$$\langle u, v \rangle_{H^\alpha(\mathcal{M}_n, \mu_n)} = \langle \mathcal{S}_{n,\alpha}^* u, \mathcal{S}_{n,\alpha}^* v \rangle_{H^\alpha(\mathcal{M}, \mu)},$$

and for all $u, v \in L^2(\mathcal{M}, \mu)$ with $\langle u, \phi_k \rangle_{L^2} = \langle v, \phi_k \rangle_{L^2} = 0$ for all $k > K(n)$ we have

$$\langle u, v \rangle_{H^\alpha(\mathcal{M}, \mu)} = \langle \mathcal{S}_{n,\alpha} u, \mathcal{S}_{n,\alpha} v \rangle_{H^\alpha(\mathcal{M}_n, \mu_n)}.$$

iii) For all $n \in \mathbb{N}$ and $k > K(n)$ the extension map $\mathcal{S}_{n,\alpha}^*$ preserves the H^α-norm for all $u \in L^2(\mathcal{M}_n, \mu_n)$ with $\langle u, \phi_k^{[n]} \rangle_{L^2(\mathcal{M}_n, \mu_n)} = 0$, and also $\mathcal{S}_{n,\alpha}$ preserves the H^α-norm for all $u \in L^2(\mathcal{M}, \mu)$ with $\langle u, \phi_k \rangle_{L^2} = 0$.

iv) It holds that $\mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} v = \sum_{k=1}^{K(n)} \langle v, \varphi_k \rangle_{L^2(\mathcal{M}, \mu)} \phi_k$ for every $v \in L^2(\mathcal{M}, \mu)$, where $\{\varphi_k\}_{k \leq K(n)}$ is any orthonormal basis of $\bigoplus_{k \leq K(n)} E(\lambda_k)$.

v) For any $v \in L^2(\mathcal{M}, \mu)$ we have $\left\| \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} v \right\|_{L^2(\mathcal{M}, \mu)} \leq \|v\|_{L^2(\mathcal{M}, \mu)}$.

Proof. Again (i), and (ii) follow immediately from the definitions and imply (iii). Furthermore (v) follows from (iv).

For (iv), writing out the definitions reads

$$
\begin{aligned}
\mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} v &= \sum_{k=1}^{K(n)} \left(\frac{1 + (\lambda_k^{[n]})^{1/2}}{1 + \lambda_k^{1/2}} \right)^\alpha \langle \mathcal{S}_{n,\alpha} v, \phi_k^{[n]} \rangle_{L^2(\mathcal{M}_n, \mu_n)} \phi_k^{(n)} \\
&= \sum_{k=1}^{K(n)} \left(\frac{1 + (\lambda_k^{[n]})^{1/2}}{1 + \sqrt{\lambda_k}} \right)^\alpha \left\langle \sum_{j=1}^{K(n)} \left(\frac{1 + \lambda_j^{1/2}}{1 + (\lambda_j^{[n]})^{1/2}} \right)^\alpha \langle v, \phi_j^{(n)} \rangle_{L^2(\mathcal{M}, \mu)} \phi_j^{[n]}, \phi_k^{[n]} \right\rangle_{L^2(\mathcal{M}_n, \mu_n)} \phi_k^{(n)} \\
&= \sum_{k=1}^{K(n)} \left(\frac{1 + (\lambda_k^{[n]})^{1/2}}{1 + \lambda_k^{1/2}} \right)^\alpha \sum_{j=1}^{K(n)} \left(\frac{1 + \lambda_k^{1/2}}{1 + (\lambda_k^{[n]})^{1/2}} \right)^\alpha \langle v, \phi_j^{(n)} \rangle_{L^2(\mathcal{M}, \mu)} \langle \phi_j^{[n]}, \phi_k^{[n]} \rangle_{L^2(\mathcal{M}_n, \mu_n)} \phi_k^{(n)} \\
&= \sum_{k=1}^{K(n)} \left(\frac{1 + (\lambda_k^{[n]})^{1/2}}{1 + \lambda_k^{1/2}} \right)^\alpha \left(\frac{1 + (\lambda_k^{[n]})^{1/2}}{1 + \lambda_k^{1/2}} \right)^{-\alpha} \langle v, \phi_k^{(n)} \rangle_{L^2(\mathcal{M}, \mu)} \phi_k^{(n)} \\
&= \sum_{k=1}^{K(n)} \langle v, \phi_k^{(n)} \rangle_{L^2(\mathcal{M}, \mu)} \phi_k^{(n)} = \mathbb{P}\left(\bigoplus_{k=1}^{K(n)} E(\lambda_k), v \right),
\end{aligned}
$$

by (4.3), which yields the claim. $\qquad \square$

So, the spectral projections project on the $K(n)$-dimensional subspaces of $L^2(\mathcal{M}_n, \mu_n)$ and $H^\alpha(\mathcal{M}_n, \mu_n)$, or $L^2(\mathcal{M}, \mu)$ and $H^\alpha(\mathcal{M}, \mu)$, respectively.

Lemma 4.4. *For any $v \in L^2(\mathcal{M}, \mu)$ it holds that $\|v - \mathcal{S}_n^* \mathcal{S}_n v\|_{L^2(\mathcal{M}, \mu)} \xrightarrow{n \to \infty} 0$, as well as $\left\| v - \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} v \right\|_{L^2(\mathcal{M}, \mu)} \xrightarrow{n \to \infty} 0$.*

Proof. Using Lemma 4.3.(iv) we observe that

$$v - \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} v = v - \mathcal{S}_n^* \mathcal{S}_n v = \sum_{k=1}^{\infty} \langle v, \varphi_k \rangle_{L^2(\mathcal{M}, \mu)} \varphi_k - \sum_{k=1}^{K(n)} \langle v, \varphi_k \rangle_{L^2(\mathcal{M}, \mu)} \varphi_k = \sum_{k=K(n)+1}^{\infty} \langle v, \varphi_k \rangle_{L^2(\mathcal{M}, \mu)} \varphi_k.$$

Thus, since $v \in L^2(\mathcal{M}, \mu)$ and $K(n) \xrightarrow{n \to \infty} \infty$ we have

$$\left\| \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} v - v \right\|_{L^2(\mathcal{M}, \mu)} = \|\mathcal{S}_n^* \mathcal{S}_n v - v\|_{L^2(\mathcal{M}, \mu)} = \left\| \sum_{k=K(n)+1}^{\infty} \langle v, \varphi_k \rangle_{L^2(\mathcal{M}, \mu)} \varphi_k \right\|_{L^2(\mathcal{M}, \mu)} \xrightarrow{n \to \infty} 0. \quad \square$$

Definition 4.5. We say a sequence of (linear) discretization operators $\mathcal{R}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$, induces graph signals consistent across resolutions, if there is a $C_{\mathcal{R}} \geq 0$ such that $\|\mathcal{R}_n u\|_{L^2} \leq C_{\mathcal{R}} \|u\|_{L^2}$ for all $n \in \mathbb{N}$ and for all $u \in L^2(\mathcal{M}, \mu)$ it holds that

$$d_{TL^2}((\mu, u), (\mu_n, \mathcal{R}_n u)) \xrightarrow{n \to \infty} 0,$$

or equivalently by (2.21)

$$\|u - \mathcal{P}_n^* u_n\|_{L^2} = \|u - \mathcal{P}_n^* \mathcal{R}_n u\|_{L^2} \xrightarrow{n \to \infty} 0. \tag{4.5}$$

In the following let $\mathcal{R}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$, $n \in \mathbb{N}$ be such a sequence of discretization operators inducing graph signals consistent across resolutions.

Remark 4.6. As can be seen from (2.21), Lemma 4.8 and Lemma 4.2 (as well as Lemma 4.10) yield that both spatial and spectral discretization are examples of such discretization operators that induce graph signals that are consistent across resolutions.

Lemma 4.7. Let $\mathcal{R}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$ satisfy Definition 4.5. Then, for any $b, v \in L^2(\mathcal{M}, \mu)$ it holds that $\|b * v - (\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n v)\|_{L^2(\mathcal{M}, \mu)} \to 0$ as $n \to \infty$.

Proof. Denote by $\{\varphi_k\}_k$ an arbitrary choice of orthonormal eigenfunctions of the Laplace-Beltrami operator. Then, bilinearity of the convolution operation and orthogonality of the eigenfunctions yield

$$
\begin{aligned}
b * v - (\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n v) &= \sum_{k=1}^{\infty} \langle b, \varphi_k \rangle \langle v, \varphi_k \rangle \varphi_k - \sum_{k=1}^{\infty} \langle \mathcal{S}_n^* \mathcal{S}_n b, \varphi_k \rangle \langle \mathcal{P}_n^* \mathcal{R}_n v, \varphi_k \rangle \varphi_k \\
&= \sum_{k=1}^{\infty} \langle b - \mathcal{S}_n^* \mathcal{S}_n b, \varphi_k \rangle \langle v, \varphi_k \rangle \varphi_k + \sum_{k=1}^{\infty} \langle \mathcal{S}_n^* \mathcal{S}_n b, \varphi_k \rangle \langle v - \mathcal{P}_n^* \mathcal{R}_n v, \varphi_k \rangle \varphi_k \\
&= \sum_{k=K(n)+1}^{\infty} \langle b, \varphi_k \rangle \langle v, \varphi_k \rangle \varphi_k + \sum_{k=1}^{K(n)} \langle b, \varphi_k \rangle \langle v - \mathcal{P}_n^* \mathcal{R}_n v, \varphi_k \rangle \varphi_k.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\|b * v - (\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n v)\|_{L^2} \\
&\qquad \leq \left\| \sum_{k=K(n)+1}^{\infty} \langle b, \varphi_k \rangle \langle v, \varphi_k \rangle \varphi_k \right\|_{L^2} + \left\| \sum_{k=1}^{K(n)} \langle b, \varphi_k \rangle \langle v - \mathcal{P}_n^* \mathcal{R}_n v, \varphi_k \rangle \varphi_k \right\|_{L^2}.
\end{aligned}
$$

Now the first term goes to 0, and for the second term it holds that

$$
\begin{aligned}
\left\| \sum_{k=1}^{K(n)} \langle b, \varphi_k \rangle \langle v - \mathcal{P}_n^* \mathcal{R}_n v, \varphi_k \rangle \varphi_k \right\|_{L^2}^2 &= \sum_{k=1}^{K(n)} \langle b, \varphi_k \rangle^2 \langle v - \mathcal{P}_n^* \mathcal{R}_n v, \varphi_k \rangle^2 \\
&\leq \sum_{k=1}^{K(n)} \langle b, \varphi_k \rangle^2 \sum_{k=1}^{K(n)} \langle v - \mathcal{P}_n^* \mathcal{R}_n v, \varphi_k \rangle^2 \\
&\leq \|\mathcal{S}_n^* \mathcal{S}_n b\|_{L^2}^2 \|v - \mathcal{P}_n^* \mathcal{R}_n v\|_{L^2}^2 \leq \|b\|_{L^2}^2 \|v - \mathcal{P}_n^* \mathcal{R}_n v\|_{L^2}^2 \xrightarrow{n \to \infty} 0,
\end{aligned}
$$

by definition of $\mathcal{R}_n$. $\qquad\square$

Lemma 4.8. For all $v \in L^2(\mathcal{M}, \mu)$ we have that $\|v - \mathcal{P}_n^* \mathcal{S}_n v\|_{L^2(\mathcal{M}, \mu)} \leq \|v\|_{L^2}^2 (\delta_n^\phi)^2 \to 0$ as $n \to \infty$.

Proof. We can expand in any $L^2(\mathcal{M}, \mu)$ orthonormal eigenbasis $\{\varphi_k\}_k$, indexed corresponding to the eigenvalues. Then, using (4.3) we can write

$$
\begin{aligned}
\|v - \mathcal{P}_n^* \mathcal{S}_n v\|_{L^2(\mathcal{M}, \mu)} &= \left\| v - \sum_{k=1}^{K(n)} \langle v, \phi_k^{(n)} \rangle \mathcal{P}_n^*(\phi_k^{[n]}) \right\|_{L^2(\mathcal{M}, \mu)} \\
&= \left\| \sum_{k=1}^{\infty} \langle v, \varphi_k \rangle \varphi_k - \sum_{k=1}^{K(n)} \langle v, \phi_k^{(n)} \rangle \mathcal{P}_n^*(\phi_k^{[n]}) \right\|_{L^2(\mathcal{M}, \mu)} \\
&= \left\| \sum_{k=1}^{K(n)} \langle v, \phi_k^{(n)} \rangle (\phi_k^{(n)} - \mathcal{P}_n^*(\phi_k^{[n]})) + \sum_{k=K(n)+1}^{\infty} \langle v, \varphi_k \rangle \varphi_k \right\|_{L^2(\mathcal{M}, \mu)} \\
&\leq \left\| \sum_{k=1}^{K(n)} |\langle v, \phi_k^{(n)} \rangle_{L^2}|(\phi_k^{(n)} - \mathcal{P}_n^*(\phi_k^{[n]})) \right\|_{L^2(\mathcal{M}, \mu)} + \|v - \mathcal{S}_n^* \mathcal{S}_n v\|_{L^2(\mathcal{M}, \mu)}.
\end{aligned}
$$

By Lemma 4.4 the last term goes to 0 as $n \to \infty$. By Lemma 3.5 we also have

$$
\begin{aligned}
\left\| \sum_{k=1}^{K(n)} |\langle v, \phi_k^{(n)} \rangle_{L^2}|(\phi_k^{(n)} - \mathcal{P}_n^*(\phi_k^{[n]})) \right\|_{L^2(\mathcal{M}, \mu)}^2 &\leq \sum_{k=1}^{K(n)} |\langle v, \phi_k^{(n)} \rangle_{L^2}|^2 \left\| \phi_k^{(n)} - \mathcal{P}_n^*(\phi_k^{[n]}) \right\|_{L^2}^2 \\
&\leq \sum_{k=1}^{K(n)} |\langle v, \phi_k^{(n)} \rangle_{L^2}|^2 (\delta_n^\phi)^2 \leq \|v\|_{L^2}^2 (\delta_n^\phi)^2 \xrightarrow{n \to \infty} 0,
\end{aligned}
$$

which yields the claim. $\qquad\square$

Lemma 4.9. *Let $\alpha \in (0, 1]$. Then, it holds for all $v \in H^\alpha(\mathcal{M}, \mu)$ that $\|\mathcal{S}_n v - \mathcal{S}_{n,\alpha} v\|_{L^2} \xrightarrow{n \to \infty} 0$.*

Proof. We have

$$
\begin{aligned}
\|\mathcal{S}_n v - \mathcal{S}_{n,\alpha} v\|_{L^2}^2 &= \left\| \sum_{k=1}^{K(n)} \left(1 - \left(\frac{1 + \lambda_k^{1/2}}{1 + (\lambda_k^{[n]})^{1/2}} \right)^\alpha \right) \langle v, \phi_k^{(n)} \rangle_{L^2} \phi_k^{[n]} \right\|_{L^2}^2 \\
&= \sum_{k=1}^{K(n)} \left| 1 - \left(\frac{1 + \lambda_k^{1/2}}{1 + (\lambda_k^{[n]})^{1/2}} \right)^\alpha \right|^2 \langle v, \phi_k^{(n)} \rangle_{L^2}^2.
\end{aligned}
$$

It holds for all $n \in \mathbb{N}$ that $\lambda_1^{[n]} = \lambda_1 = 0$, and since $\mathcal{M}$ is connected $\lambda_k > 0$ for all $k \geq 2$. Now we can further estimate using $|x + y|^\alpha \leq |x|^\alpha + |y|^\alpha$ (and as a consequence $\||x|^\alpha - |y|^\alpha| \leq |x - y|^\alpha$) for all $x, y \in \mathbb{R}$ and $\alpha \in (0, 1]$ and Theorem 3.1, in particular (3.1), that for all $2 \leq k \leq n$

$$
\begin{aligned}
\left| 1^\alpha - \left(\frac{1 + \lambda_k^{1/2}}{1 + (\lambda_k^{[n]})^{1/2}} \right)^\alpha \right|^2 &\leq \left| 1 - \frac{1 + \lambda_k^{1/2}}{1 + (\lambda_k^{[n]})^{1/2}} \right|^{2\alpha} = \left| \frac{(\lambda_k^{[n]})^{1/2} - \lambda_k^{1/2}}{1 + (\lambda_k^{[n]})^{1/2}} \right|^{2\alpha} \\
&\leq \left(\frac{\left| \lambda_k^{[n]} - \lambda_k \right|^{1/2}}{\left| 1 + (\lambda_k^{[n]}) \right|^{1/2}} \right)^{2\alpha} = \left| \frac{\lambda_k^{[n]} - \lambda_k}{1 + (\lambda_k^{[n]})} \right|^\alpha \leq \left| \lambda_k^{[n]} - \lambda_k \right|^\alpha \\
&= \left(\frac{\left| \lambda_k^{[n]} - \lambda_k \right|}{1 + \lambda_k} \right)^\alpha (1 + \lambda_k)^\alpha \leq \left(\frac{\left| \lambda_k^{[n]} - \lambda_k \right|}{\lambda_k} \right)^\alpha (1 + \lambda_k)^\alpha \\
&\leq (C\delta(\varepsilon_n, h_n, \lambda_k))^\alpha (1 + \lambda_k)^\alpha \leq C\delta(\varepsilon_n, h_n, \lambda_{K(n)})^\alpha (1 + \lambda_k)^\alpha.
\end{aligned}
$$

25

Putting this together with the above we have for all $n \in \mathbb{N}$

$$\|\mathcal{S}_n v - \mathcal{S}_{n,\alpha} v\|_{L^2}^2 = \sum_{k=1}^{K(n)} \left| 1 - \left(\frac{1 + \lambda_k^{1/2}}{1 + (\lambda_k^{[n]})^{1/2}} \right)^\alpha \right|^2 \langle v, \phi_k^{(n)} \rangle_{L^2}^2$$

$$= \langle v, \phi_1^{(n)} \rangle_{L^2}^2 + \sum_{k=2}^{K(n)} C \left(\delta(\varepsilon_n, h_n, \lambda_{K(n)}) \right)^\alpha (1 + \lambda_k)^\alpha \langle v, \phi_k^{(n)} \rangle_{L^2}^2$$

$$\leq C \delta(\varepsilon_n, h_n, \lambda_{K(n)})^\alpha \sum_{k=1}^{K(n)} (1 + \lambda_k)^\alpha \langle v, \phi_k^{(n)} \rangle_{L^2}^2$$

$$\leq C \delta(\varepsilon_n, h_n, \lambda_{K(n)})^\alpha \|v\|_{H^\alpha}^2 .$$

So (3.10) yields that this converges to 0 as $n \to \infty$. $\qquad \square$

Lemma 4.10. *It holds for all $v \in H^\alpha(\mathcal{M}, \mu)$, that $\left\| \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} v - \mathcal{P}_n^* \mathcal{S}_{n,\alpha} v \right\|_{L^2(\mathcal{M},\mu)} \to 0$ as $n \to \infty$.*

Proof. Using Lemma 4.3.(iv) and Lemma 4.1.(iii) we obtain

$$\left\| \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} v - \mathcal{P}_n^* \mathcal{S}_{n,\alpha} v \right\|_{L^2(\mathcal{M},\mu)} = \left\| \mathcal{S}_n^* \mathcal{S}_n v - \mathcal{P}_n^* \mathcal{S}_{n,\alpha} v \right\|_{L^2(\mathcal{M},\mu)}$$

$$\leq \left\| \mathcal{S}_n^* \mathcal{S}_n v - \mathcal{P}_n^* \mathcal{S}_n v \right\|_{L^2(\mathcal{M},\mu)} + \left\| \mathcal{P}_n^* \mathcal{S}_n v - \mathcal{P}_n^* \mathcal{S}_{n,\alpha} v \right\|_{L^2(\mathcal{M},\mu)}$$

$$= \left\| \mathcal{S}_n^* \mathcal{S}_n v - \mathcal{P}_n^* \mathcal{S}_n v \right\|_{L^2(\mathcal{M},\mu)} + \left\| \mathcal{S}_n v - \mathcal{S}_{n,\alpha} v \right\|_{L^2(\mathcal{M}_n, \mu_n)} .$$

where both terms go to 0 as $n \to \infty$ by Lemma 4.8 and Lemma 4.9. $\qquad \square$

With the preceding convergence results for the discretization and extension maps, we can now compare a lifted discrete and a continuum convolution expression with appropriate discretizations of the convolution parameter b and the signal. Using a spectral discretization on b is essential, since it allows us to apply the uniform error bounds on the graph-eigenfunctions from Lemma 3.5. This yields the following estimate:

Lemma 4.11. *Let $\mathcal{R}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$ satisfy Definition 4.5. Then, for any $b, v \in L^2(\mathcal{M}, \mu)$ it holds that*

$$\|\mathcal{P}_n^*(\mathcal{S}_n b *_n \mathcal{R}_n v) - (\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n v)\|_{L^2(\mathcal{M},\mu)} \leq 2 C_{\mathcal{R}_n} \delta_n^\phi \|b\|_{L^2} \|u\|_{L^2} \xrightarrow{n \to \infty} 0.$$

Proof. Again we expand in an arbitrary $L^2(\mathcal{M}, \mu)$ orthonormal eigenbasis $\{\varphi_k\}_k$, indexed corresponding to the eigenvalues. Using Lemma 4.1.(ii), the extended discrete convolution expression reads

$$\mathcal{P}_n^*(\mathcal{S}_n b *_n \mathcal{R}_n v) = \sum_{k=1}^{K(n)} \langle \mathcal{S}_n b, \phi_k^{[n]} \rangle_{L^2(\mathcal{M}_n, \mu_n)} \langle \phi_k^{[n]}, \mathcal{R}_n v \rangle_{L^2(\mathcal{M}_n, \mu_n)} \mathcal{P}_n^* \phi_k^{[n]}$$

$$= \sum_{k=1}^{K(n)} \langle b, \phi_k^{(n)} \rangle_{L^2(\mathcal{M},\mu)} \langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} \rangle_{L^2(\mathcal{M},\mu)} \mathcal{P}_n^* \phi_k^{[n]},$$

while using (4.3), the continuum convolution expression reads

$$(\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n v) = \sum_{k=1}^{\infty} \langle \mathcal{S}_n^* \mathcal{S}_n b, \varphi_k \rangle_{L^2(\mathcal{M},\mu)} \langle \mathcal{P}_n^* \mathcal{R}_n v, \varphi_k \rangle_{L^2(\mathcal{M},\mu)} \varphi_k$$

$$= \sum_{k=1}^{K(n)} \langle b, \varphi_k \rangle_{L^2(\mathcal{M},\mu)} \langle \mathcal{P}_n^* \mathcal{R}_n v, \varphi_k \rangle_{L^2(\mathcal{M},\mu)} \varphi_k$$

$$= \sum_{k=1}^{K(n)} \langle b, \phi_k^{(n)} \rangle_{L^2(\mathcal{M},\mu)} \langle \mathcal{P}_n^* \mathcal{R}_n v, \phi_k^{(n)} \rangle_{L^2(\mathcal{M},\mu)} \phi_k^{(n)}.$$

Therefore

$$\|\mathcal{P}_n^*(\mathcal{S}_n b *_n \mathcal{R}_n v) - (\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n v)\|_{L^2}$$

$$= \left\| \sum_{k=1}^{K(n)} \langle b, \phi_k^{(n)} \rangle_{L^2} \left(\langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} \rangle_{L^2} \mathcal{P}_n^* \phi_k^{[n]} - \langle \mathcal{P}_n^* \mathcal{R}_n v, \phi_k^{(n)} \rangle_{L^2} \phi_k^{(n)} \right) \right\|_{L^2}$$

$$= \left\| \sum_{k=1}^{K(n)} \langle b, \phi_k^{(n)} \rangle_{L^2} \left(\langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)} \rangle_{L^2} \phi_k^{(n)} + \langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} \rangle_{L^2} (\mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)}) \right) \right\|_{L^2}$$

$$\leq \left\| \sum_{k=1}^{K(n)} \langle b, \phi_k^{(n)} \rangle_{L^2} \langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)} \rangle_{L^2(\mathcal{M},\mu)} \phi_k^{(n)} \right\|_{L^2}$$

$$+ \left\| \sum_{k=1}^{K(n)} \langle b, \phi_k^{(n)} \rangle_{L^2} \langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} \rangle_{L^2} (\mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)}) \right\|_{L^2}.$$

Now for the first term using Lemma 3.5, Lemma 4.3.(v) and Definition 4.5 we get that

$$\left\| \sum_{k=1}^{K(n)} \langle b, \phi_k^{(n)} \rangle_{L^2} \langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)} \rangle_{L^2} \phi_k \right\|_{L^2}^2 = \sum_{k=1}^{K(n)} \langle b, \phi_k \rangle_{L^2}^2 \langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)} \rangle_{L^2}^2$$

$$\leq \sum_{k=1}^{K(n)} \langle b, \phi_k^{(n)} \rangle_{L^2}^2 \| \mathcal{P}_n^* \mathcal{R}_n v \|_{L^2}^2 \left\| \mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)} \right\|_{L^2}^2$$

$$\leq \| \mathcal{P}_n^* \mathcal{R}_n v \|_{L^2}^2 (\delta_n^\phi)^2 \| \mathcal{S}_n^* \mathcal{S}_n b \|_{L^2}^2$$

$$\leq C_{\mathcal{R}} \| v \|_{L^2}^2 (\delta_n^\phi)^2 \| b \|_{L^2}^2 \xrightarrow{n \to \infty} 0,$$

while for the second term we get using the same results that

$$\left\| \sum_{k=1}^{K(n)} \langle b, \phi_k^{(n)} \rangle_{L^2} \langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} \rangle_{L^2} (\mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)}) \right\|_{L^2}$$

$$\leq \sum_{k=1}^{K(n)} |\langle b, \phi_k^{(n)} \rangle_{L^2}| |\langle \mathcal{P}_n^* \mathcal{R}_n v, \mathcal{P}_n^* \phi_k^{[n]} \rangle_{L^2}| \left\| \mathcal{P}_n^* \phi_k^{[n]} - \phi_k^{(n)} \right\|_{L^2}$$

$$\leq \delta_n^\phi \sum_{k=1}^{K(n)} |\langle b, \phi_k^{(n)} \rangle_{L^2}| |\langle \mathcal{R}_n v, \phi_k^{[n]} \rangle_{L^2(\mathcal{M}_n,\mu_n)}|$$

$$\leq \delta_n^\phi \| \mathcal{S}_n^* \mathcal{S}_n b \|_{L^2(\mathcal{M},\mu)} \| \mathcal{R}_n v \|_{L^2(\mathcal{M}_n,\mu_n)}$$

$$\leq C_{\mathcal{R}} \delta_n^\phi \| b \|_{L^2} \| v \|_{L^2} \xrightarrow{n \to \infty} 0.$$

So, with all the above, we obtain

$$\|\mathcal{P}_n^*(\mathcal{S}_n b *_n \mathcal{R}_n v) - (\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n v)\|_{L^2} \leq 2 C_{\mathcal{R}} \delta_n^\phi \| b \|_{L^2} \| v \|_{L^2} \xrightarrow{n \to \infty} 0. \qquad \square$$

4.3 Interplay with neural response maps

These discretization maps can be naturally used in constructing discretization maps and extension maps between the (ambient) continuum parameter space Θ and the (ambient) discrete parameter spaces Θ_n, $n \in \mathbb{N}$, respectively vice versa:

$$\mathcal{Q}_{n,\alpha} : (L^2(\mathcal{M},\mu))^3 \to (L^2(\mathcal{M}_n,\mu_n))^3, \quad (a,b,c) \mapsto (\mathcal{S}_{n,\alpha} a, \mathcal{S}_n b, \mathcal{S}_{n,\alpha} c), \tag{4.6}$$

27

and

$$\mathcal{Q}_{n,\alpha}^* : (L^2(\mathcal{M}_n, \mu_n))^3 \to (L^2(\mathcal{M}, \mu))^3, \quad (a, b, c) \mapsto (\mathcal{S}_{n,\alpha}^* a, \mathcal{S}_n^* b, \mathcal{S}_{n,\alpha}^* c). \tag{4.7}$$

Remark 4.12. So we observe that the discretization and extension maps just map the continuum ambient space to the subspace spanned by the first $K(n)$ eigenvectors of the n-th graph Laplacian of the respective discrete space and vice versa by Lemma 4.3.(iv) (which we denote by adding the subscript $K(n)$ to a space), i.e. we have that:

$$\mathcal{Q}_{n,\alpha}(H^\alpha(\mathcal{M}) \times L^2(\mathcal{M}) \times H^\alpha(\mathcal{M})) = H^\alpha_{K(n)}(\mathcal{M}_n) \times L^2_{K(n)}(\mathcal{M}_n) \times H^\alpha_{K(n)}(\mathcal{M}_n), \tag{4.8}$$

$$\mathcal{Q}_{n,\alpha}^*(H^\alpha(\mathcal{M}_n) \times L^2(\mathcal{M}_n) \times H^\alpha(\mathcal{M}_n)) = H^\alpha_{K(n)}(\mathcal{M}) \times L^2_{K(n)}(\mathcal{M}) \times H^\alpha_{K(n)}(\mathcal{M}), \tag{4.9}$$

and putting them together yields the relations

$$\mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha}(H^\alpha(\mathcal{M}) \times L^2(\mathcal{M}) \times H^\alpha(\mathcal{M})) = H^\alpha_{K(n)}(\mathcal{M}) \times L^2_{K(n)}(\mathcal{M}) \times H^\alpha_{K(n)}(\mathcal{M}), \tag{4.10}$$

$$\mathcal{Q}_{n,\alpha}(H^\alpha_{K(n)}(\mathcal{M}) \times L^2_{K(n)}(\mathcal{M}) \times H^\alpha_{K(n)}(\mathcal{M})) = H^\alpha_{K(n)}(\mathcal{M}_n) \times L^2_{K(n)}(\mathcal{M}_n) \times H^\alpha_{K(n)}(\mathcal{M}_n), \tag{4.11}$$

More precisely, taking norm preservation into account we get

$$\mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha}(\Theta) = \left(B_{H^\alpha} \cap H^\alpha_{K(n)}(\mathcal{M})\right) \times \left(B_{L^2} \cap L^2_{K(n)}(\mathcal{M})\right) \times \left(B_{H^\alpha} \cap H^\alpha_{K(n)}(\mathcal{M})\right), \tag{4.12}$$

and

$$\mathcal{Q}_{n,\alpha} \mathcal{Q}_{n,\alpha}^*(\Theta_n) = \Theta_n. \tag{4.13}$$

So we have that $\mathcal{Q}_{n,\alpha}^*$ and $\mathcal{Q}_{n,\alpha}$ are homeomorphisms between the $K(n)$-dimensional subspaces of the parameter spaces equipped with their respective topology, obtained by the spectral cutoff of the respective ambient parameter spaces $H^\alpha \times L^2 \times H^\alpha$:

Lemma 4.13. *For the parameter discretization and extension maps it holds that $\mathcal{Q}_{n,\alpha}(\Theta) = \Theta_n$ and*

$$\mathcal{Q}_{n,\alpha}^* : \Theta_n \to (\mathcal{Q}_{n,\alpha}^*(\Theta_n), w)$$

is an homeomorphism with the inverse being $\mathcal{Q}_{n,\alpha} : (\mathcal{Q}_{n,\alpha}^*(\Theta_n), w) \to \Theta_n$.

Proof. The first claim follows from the observation (4.11) that the subspaces spanned by the first $K(n)$ eigenvectors or eigenfunctions agree under the respective projection due to Lemma 4.3.(iv). Regarding the second claim, first note that $\mathcal{Q}_{n,\alpha}$ and its adjoint $\mathcal{Q}_{n,\alpha}^*$ are linear bounded operators. Furthermore using (4.13) we see that $\mathcal{Q}_{n,\alpha}^* : \Theta_n \to (\mathcal{Q}_{n,\alpha}^*(\Theta_n), w)$ is bijective with inverse $\mathcal{Q}_{n,\alpha} : (\mathcal{Q}_{n,\alpha}^*(\Theta_n), w) \to \Theta_n$. Finally $\mathcal{Q}_{n,\alpha}^*$ and $\mathcal{Q}_{n,\alpha}$ are continuous w.r.t the appropriate topologies, because as bounded linear operators they are norm-norm-continuous, but also weak-weak-continuous, which yields the second claim. $\qquad \square$

Note that the discretization and extension maps restricted to Θ_n and $\mathcal{Q}_{n,\alpha}\Theta_n$ are L^2-product norm preserving due to Lemma 4.3.(iii).

Remark 4.14. An immediate consequence of Lemma 4.13 is that

$$C(\mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha}\Theta, w) = C(\mathcal{Q}_{n,\alpha}\Theta_n) \cong C(\Theta_n) \tag{4.14}$$

and

$$\mathcal{M}(\mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha}\Theta) = \mathcal{M}(\mathcal{Q}_{n,\alpha}^* \Theta_n) = (\mathcal{Q}_{n,\alpha}^*)_\# \mathcal{M}(\Theta_n) \cong \mathcal{M}(\Theta_n). \tag{4.15}$$

Remark 4.15. For $u, u' \in L^2(\mathcal{M}, \mu)$ and $\theta, \theta' \in \left(L^2(\mathcal{M}, \mu) \right)^3$, where we write $\theta = (a, b, c)$ and $\theta' = (a', b', c')$, we get using Cauchy-Schwarz, (2.9) and (2.10)

$$
\begin{aligned}
|\psi(u, \theta) - \psi(u', \theta')| &= \left| \langle a, \sigma(b * u + c) \rangle - \langle a', \sigma(b' * u' + c') \rangle \right| \\
&\leq \left\| (a - a') \right\|_{L^2} C_\sigma \left(1 + \|b * u\|_{L^2} + \|c\|_{L^2} \right) \\
&\quad + \|a'\|_{L^2} L_\sigma \left(\|b * u - b' * u'\|_{L^2} + \|c - c'\|_{L^2} \right).
\end{aligned}
$$

Lemma 4.16. *Let* $\mathcal{R}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$ *satisfy Definition 4.5,* $u \in L^2(\mathcal{M}, \mu)$ *and* $\theta \in \Theta$. *Then it holds that*

$$
|\psi(\mathcal{P}_n^* \mathcal{R}_n u, \mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha} \theta) - \psi(u, \theta)| \xrightarrow{n \to \infty} 0.
$$

Proof. Remark 4.15 yields

$$
\begin{aligned}
|\psi(u, \theta) - \psi(\mathcal{P}_n^* \mathcal{R}_n u, \mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha} \theta)| &\leq \left\| (a - \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} a) \right\|_{L^2} C_\sigma \left(1 + \|b * u\|_{L^2} + \|c\|_{L^2} \right) \\
&\quad + \left\| \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} a - a \right\|_{L^2} L_\sigma \left(\|b * u - \mathcal{S}_n^* \mathcal{S}_n b * \mathcal{P}_n^* \mathcal{R}_n u\|_{L^2} \right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left. + \left\| c - \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c \right\|_{L^2} \right) \\
&\quad + \|a\|_{L^2} L_\sigma \left(\|b * u - \mathcal{S}_n^* \mathcal{S}_n b * \mathcal{P}_n^* \mathcal{R}_n u\|_{L^2} + \left\| c - \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c \right\|_{L^2} \right),
\end{aligned}
$$

for which Lemma 4.4 and Lemma 4.7 yield that the right-hand-side above goes to 0 as $n \to \infty$. $\quad\square$

Lemma 4.17. *Let* $\mathcal{R}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$ *satisfy Definition 4.5,* $u \in L^2(\mathcal{M}, \mu)$ *and* $\theta \in \Theta$. *Then it holds that*

$$
|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha} \theta) - \psi(\mathcal{P}_n^* \mathcal{R}_n u, \mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha} \theta)| \xrightarrow{n \to \infty} 0.
$$

Proof. Let $a, b, c \in L^2(\mathcal{M}, \mu)$, $\|a\|_{L^2} \leq 1$, $\|b\|_{L^2} \leq 1$, $\|c\|_{L^2} \leq 1$ such that $\theta = (a, b, c)$.

We observe using Lemma 4.1.(ii), Cauchy-Schwarz and Equation (2.10) that

$$
\begin{aligned}
&|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha} \theta) - \psi(\mathcal{P}_n^* \mathcal{R}_n u, \mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha} \theta)| && (4.16) \\
&= |\langle \mathcal{S}_{n,\alpha} a, \sigma(\mathcal{S}_n b *_n \mathcal{R}_n u + \mathcal{S}_{n,\alpha} c) \rangle_{L^2(\mathcal{M}_n, \mu_n)} \\
&\qquad - \langle \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} a, \sigma((\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) + \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c) \rangle_{L^2(\mathcal{M}, \mu)}| \\
&= |\langle \mathcal{P}_n^* \mathcal{S}_{n,\alpha} a, \mathcal{P}_n^* \big(\sigma(\mathcal{S}_n b *_n \mathcal{R}_n u + \mathcal{S}_{n,\alpha} c) \big) - \sigma((\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) + \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c) \rangle_{L^2(\mathcal{M}, \mu)} \\
&\qquad + \langle \mathcal{P}_n^* \mathcal{S}_{n,\alpha} a - \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} a, \sigma((\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) + \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c) \rangle_{L^2(\mathcal{M}, \mu)}| \\
&\leq \|\mathcal{P}_n^* \mathcal{S}_{n,\alpha} a\|_{L^2} \|\mathcal{P}_n^* \big(\sigma(\mathcal{S}_n b *_n \mathcal{R}_n u + \mathcal{S}_{n,\alpha} c) \big) - \sigma((\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) + \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c)\|_{L^2} \\
&\qquad + \left\| \mathcal{P}_n^* \mathcal{S}_{n,\alpha} a - \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} a \right\|_{L^2} C_\sigma \left(\|(\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) + \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c\|_{L^2} + 1 \right).
\end{aligned}
$$

By Lemma 4.10 it holds that

$$
\|\mathcal{P}_n^* \mathcal{S}_{n,\alpha} a\|_{L^2} \leq \left\| \mathcal{P}_n^* \mathcal{S}_{n,\alpha} a - \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} a \right\|_{L^2} + \left\| \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} a - a \right\|_{L^2} + \|a\|_{L^2} \xrightarrow{n \to \infty} \|a\|_{L^2},
$$

and additionally Lemma 4.7 and Lemma 4.4

$$
\|\mathcal{P}_n^* \mathcal{S}_{n,\alpha} a - \mathcal{S}_n^* \mathcal{S}_{n,\alpha} a\|_{L^2} C_\sigma \left(\|(\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) + \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c\|_{L^2} + 1 \right) \xrightarrow{n \to \infty} 0,
$$

so it remains to treat the other term.

With Lemma 4.1.(vi), the linearity of $\mathcal{P}_n^*$, and the Lipschitz property of σ, we further obtain

$$
\left\| \mathcal{P}_n^* \big(\sigma(\mathcal{S}_n b *_n \mathcal{R}_n u + \mathcal{S}_{n,\alpha} c) \big) - \sigma((\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) + \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c) \right\|_{L^2}
$$

$$= \left\| \sigma\big(\mathcal{P}_n^*(\mathcal{S}_n b *_n \mathcal{R}_n u + \mathcal{S}_{n,\alpha} c)\big) - \sigma\big((\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) + \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c\big) \right\|_{L^2}$$

$$\leq L_\sigma \left\| \mathcal{P}_n^*(\mathcal{S}_n b *_n \mathcal{R}_n u + \mathcal{S}_{n,\alpha} c) - \big((\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) + \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c\big) \right\|_{L^2}$$

$$\leq L_\sigma \left(\left\| \mathcal{P}_n^*(\mathcal{S}_n b *_n \mathcal{R}_n u) - (\mathcal{S}_n^* \mathcal{S}_n b) * (\mathcal{P}_n^* \mathcal{R}_n u) \right\|_{L^2} + \left\| \mathcal{P}_n^* \mathcal{S}_{n,\alpha} c - \mathcal{S}_{n,\alpha}^* \mathcal{S}_{n,\alpha} c \right\| \right),$$

which goes to 0 as $n \to \infty$ by Lemma 4.11 and Lemma 4.10. Hence, we conclude that

$$\left| \psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha} \theta) - \psi(\mathcal{P}_n^* \mathcal{R}_n u, \mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha} \theta) \right| \xrightarrow{n \to \infty} 0. \qquad \square$$

As a corollary of the past two lemmas Lemma 4.16 and Lemma 4.17 we obtain immediately:

Corollary 4.18. *Let $\mathcal{R}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$ satisfy Definition 4.5, $u \in L^2(\mathcal{M}, \mu)$ and $\theta \in \Theta$. Then it holds that*

$$\left| \psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha} \theta) - \psi(u, \theta) \right| \xrightarrow{n \to \infty} 0.$$

5 Γ-Convergence of the ERM functionals

Let $\{(u_k, y_k)\}_{k=1}^l \in (L^2(\mathcal{M}) \times \mathbb{R})^l$ be continuum training data, coming from a functional on the space of manifold signals $g : L^2(\mathcal{M}, \mu) \to \mathbb{R}$, so $g(u_k) = y_k$.

From these, using a sequence of discretization operators $\mathcal{R}_n : L^2(\mathcal{M}, \mu) \to L^2(\mathcal{M}_n, \mu_n)$ satisfying Definition 4.5, we obtain the discrete training data $\{(\mathcal{R}_n u_k, y_k)\}_{k=1}^l \in (L^2(\mathcal{M}_n, \mu_n) \times \mathbb{R})^l$, and the underlying functionals $g_n : L^2(\mathcal{M}_n, \mu_n) \to \mathbb{R}$ on the space of graph signals such that $g_n(\mathcal{R}_n u_k) = y_k = g(u_k)$.

Then, given a non-negative real-valued continuous loss function $\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$, and a regularization parameter $\zeta \geq 0$ the (regularized) continuum ERM functional $J_\alpha : \mathcal{M}(\Theta) \to \mathbb{R}^+ \cup \{+\infty\}$ for the representation defined in (2.12) is given by:

$$J_\alpha(\nu) := \frac{1}{l} \sum_{k=1}^l \ell(f_\nu(u_k), y_k) + \zeta \|\nu\|_{TV} \qquad \text{for } \nu \in \mathcal{M}(\Theta). \tag{5.1}$$

Analogously, for $n \in \mathbb{N}$ and discrete training data as defined above, we define the (regularized) ERM functionals $\tilde{J}_{n,\alpha} : \mathcal{M}(\Theta_n) \to \mathbb{R}^+ \cup \{+\infty\}$ for the discrete representation of (2.11) as follows:

$$\tilde{J}_{n,\alpha}(\tilde{\nu}) := \frac{1}{l} \sum_{k=1}^l \ell(f_{\tilde{\nu}}^{[n]}(\mathcal{R}_n u_k), y_k) + \zeta \|\tilde{\nu}\|_{TV} \qquad \text{for } \tilde{\nu} \in \mathcal{M}(\Theta_n). \tag{5.2}$$

Then, define $J_{n,\alpha} : \mathcal{M}(\Theta) \to \mathbb{R}^+ \cup \{+\infty\}$ as

$$J_{n,\alpha}(\nu) = \begin{cases} \frac{1}{l} \sum_{k=1}^l \ell(f_{(\mathcal{Q}_{n,\alpha})_\# \nu}^{[n]}(\mathcal{R}_n u_k), y_k) + \zeta \left\| (\mathcal{Q}_{n,\alpha})_\# \nu \right\|_{TV}, & (\mathcal{Q}_{n,\alpha}^*)_\# (\mathcal{Q}_{n,\alpha})_\# \nu = \nu, \\ +\infty, & \text{otherwise.} \end{cases} \tag{5.3}$$

Notice that

$$J_{n,\alpha}(\nu) = \begin{cases} \tilde{J}_{n,\alpha}((\mathcal{Q}_{n,\alpha})_\# \nu) & (\mathcal{Q}_{n,\alpha}^*)_\# (\mathcal{Q}_{n,\alpha})_\# \nu = \nu, \\ +\infty, & \text{otherwise,} \end{cases}$$

so because of that and (4.15) we see that minimizing $J_{n,\alpha}$ and $\tilde{J}_{n,\alpha}$ is equivalent. Moreover, in the setting under consideration of supervised learning with a loss involving the finite number l of data points, all of these functionals admit discrete measures as minimizers.

Proposition 5.1. *Assume that $\zeta > 0$ and that for any $y \in \mathbb{R}$, the function $z \mapsto \ell(z, y)$ is convex and coercive. Then, the functionals $J_\alpha, \tilde{J}_{n,\alpha}$, and $J_{n,\alpha}$ admit at least one minimizer supported on at most l points of Θ, Θ_n and $\mathcal{Q}_{n,\alpha}^* \Theta_n$, respectively.*

Proof. We aim to use the general representer theorem of [5, Thm. 3.3]. Now, we notice that

$$\left\{\nu \in \mathcal{M}(\Theta) \,\middle|\, (\mathcal{Q}_{n,\alpha}^*)_\# (\mathcal{Q}_{n,\alpha})_\# \nu = \nu\right\} = \mathcal{Q}_{n,\alpha}^*\big(\mathcal{M}(\Theta_n)\big), \tag{5.4}$$

as can be seen directly from $\Theta_n = \mathcal{Q}_{n,\alpha}\Theta$. Moreover, by continuity of $\mathcal{Q}_{n,\alpha}$ and $\mathcal{Q}_{n,\alpha}^*$, the maps $(\mathcal{Q}_{n,\alpha})_\#$ and $(\mathcal{Q}_{n,\alpha}^*)_\#$ preserve weak* convergence of measures, so the set in (5.4) is a weak* closed subspace of $\mathcal{M}(\Theta)$. Therefore, taking into account the assumptions on ℓ, the assumptions of [5, Thm. 3.3] are satisfied in all three cases. For J_α and $\tilde{J}_{n,\alpha}$ the regularizer is the total variation norm for spaces of measures, with the extreme points of its sub level-sets being proportional to Dirac masses [48, Thm. 2]. For $J_{n,\alpha}$ we can use that the map $(\mathcal{Q}_{n,\alpha})_\#$ is linear and injective on the set (5.4), which implies [5, Lem. 3.2] that the extreme points of the image with respect to this map are the images of the extreme points. $\qquad\square$

Now, we will show that for any regularization parameter $\zeta \geq 0$ the functionals $J_{n,\alpha}$ Γ-converge to J_α, i.e $J_{n,\alpha} \xrightarrow{\Gamma} J_\alpha$ on the space of finite Radon measures, equipped with the weak-* topology $(\mathcal{M}(\Theta), w^*)$.

Theorem 5.2 (Uniform convergence of the neural response for consistent data)**.** *For any given* $u \in L^2(\mathcal{M}, \mu)$, *the sequence of maps*

$$\{\psi_n(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_n - \psi(u, \cdot)\}_n \subset C((\Theta, d_w))$$

is relatively compact in the topology induced by the uniform norm, where d_w *denotes a metric metrizing the weak topology.*

Remark 5.3. The weak topology on the unit ball of $H^\alpha(\mathcal{M}, \mu)$ and $L^2(\mathcal{M}, \mu)$ is metrizable, since they are both separable Hilbert spaces [6, Thm. 3.28]. Furthermore the weak topology of a finite product equipped with the maximum metric clearly coincides with the product of the spaces with their respective weak topology, so Θ is also metrizable.
Rellich-Kondrachov for fractional Sobolev spaces [36, Lem. 10] implies that weak H^α convergence implies a strongly L^2-converging subsequence. On bounded subsets of L^2 it also holds, that strong L^2 convergence implies weak H^α convergence up to a subsequence, because every bounded sequence has a weakly convergent subsequence. Since limits are unique, and the L^2-norm is bounded by the H^α-norm, this implies that

$$(\Theta, w) \to B_{L^2} \times B_{L^2}, \quad (a, b, c) \mapsto (a, c)$$

is weak-strong continuous in a and c.

Proof of Theorem 5.2. First, we will show that for all $u \in L^2(\mathcal{M}, \mu)$, $\theta \in \Theta$ and $\varepsilon > 0$ there is a $N_\varepsilon \in \mathbb{N}$ and $\delta > 0$ such that for all $\theta' \in B_{d_w}(\theta, \delta) \cap \Theta := \{\vartheta \in \Theta \mid d_w(\vartheta, \theta) < \delta\}$ and $n \geq N_\varepsilon$ it holds that $|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta) - \psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta')| < \varepsilon$.

Let $u \in L^2(\mathcal{M}, \mu)$ be arbitrary but fixed. Fix a $\theta = (a, b, c) \in \Theta$ and $\varepsilon > 0$. Let

$$d := \frac{\varepsilon}{C_\sigma\left(2 + C_\mathcal{R}\|u\|_{L^2}\right) + 5L_\sigma}.$$

The compactness of the convolution operator $(\cdot * u) \in L(L^2(\mathcal{M}, \mu), L^2(\mathcal{M}, \mu))$ yields that for any $v \in L^2(\mathcal{M}, \mu)$ and any strong L^2-neighborhood V of $v * u$ there exists a weak L^2-neighborhood N such that $\{v' * u \mid v' \in N\} =: N * u \subset V$. Following Remark 5.3, we can pick a $0 < \delta$ small enough such that for all $\theta' = (a', b', c')$ with $d_w(\theta, \theta') < \delta$ we have $\|a - a'\|_{L^2(\mathcal{M}, \mu)} < d$, $\|(b - b') * u\|_{L^2} < d$ and $\|c - c'\|_{L^2(\mathcal{M}, \mu)} < d$.

Furthermore Lemma 4.11 and the proof of Lemma 4.7 yield that there exists a $N_\varepsilon \in \mathbb{N}$ such that for all $n \geq N_\varepsilon$ and $\tilde{b} \in L^2(\mathcal{M}, \mu)$ it is

$$
\begin{aligned}
\left\|\mathcal{S}_n\tilde{b} *_n \mathcal{R}_n u\right\|_{L^2(\mathcal{M}_n,\mu_n)} &= \left\|\mathcal{P}_n^*(\mathcal{S}_n\tilde{b} *_n \mathcal{R}_n u)\right\|_{L^2(\mathcal{M},\mu)} \\
&\leq \left\|\tilde{b} * u\right\|_{L^2(\mathcal{M},\mu)} + \left\|\mathcal{P}_n^*(\mathcal{S}_n\tilde{b} *_n \mathcal{R}_n u) - \mathcal{S}_n^*\mathcal{S}_n\tilde{b} * \mathcal{P}_n^*\mathcal{R}_n u\right\|_{L^2} \\
&\quad + \left\|\mathcal{S}_n^*\mathcal{S}_n\tilde{b} * \mathcal{P}_n^*\mathcal{R}_n u - \tilde{b} * u\right\|_{L^2} \\
&\leq 2\left\|\tilde{b} * u\right\|_{L^2(\mathcal{M},\mu)} + \left\|\tilde{b}\right\|_{L^2}\left(2C_\mathcal{R}\delta_n^\phi\left\|u\right\|_{L^2} + \left\|u - \mathcal{P}_n^*\mathcal{R}_n u\right\|_{L^2}\right) \\
&\leq 2\left\|v * u\right\|_{L^2(\mathcal{M},\mu)} + \left\|v\right\|_{L^2} d,
\end{aligned}
$$

where the first equality comes from Lemma 4.1.(iii). In the following calculation we are using again the bilinear extension argument, the Lipschitz property of σ, the linearity of the discretization operators, the contractiveness regarding the L^2-norm of the spatial and spectral discretization operators, and the fact that Θ is defined as the product of H^α and L^2 unit balls and is therefore contained in the product L^2 unit balls.

For all $n \geq N_\varepsilon$ and any $\theta' = (a', b', c') \in \Theta$, such that $d_w(\theta, \theta') < \delta$ it holds that

$$
\begin{aligned}
&\left|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta) - \psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta')\right| \\
&= \left|\langle\mathcal{S}_{n,\alpha}a, \sigma(\mathcal{S}_n b *_n \mathcal{R}_n u + \mathcal{S}_{n,\alpha}c)\rangle_{L^2(\mathcal{M}_n)} - \langle\mathcal{S}_{n,\alpha}a', \sigma(\mathcal{S}_n b' *_n \mathcal{R}_n u + \mathcal{S}_{n,\alpha}c')\rangle_{L^2(\mathcal{M}_n)}\right| \\
&\leq \left\|\mathcal{S}_{n,\alpha}(a - a')\right\|_{L^2(\mathcal{M}_n)} C_\sigma\left(1 + \left\|\mathcal{S}_n b' *_n \mathcal{R}_n u\right\|_{L^2(\mathcal{M}_n)} + \left\|\mathcal{S}_{n,\alpha}c'\right\|_{L^2(\mathcal{M}_n)}\right) \\
&\quad + \left\|\mathcal{S}_{n,\alpha}a\right\|_{L^2(\mathcal{M}_n)} L_\sigma\left(\left\|(\mathcal{S}_n(b - b')) *_n \mathcal{R}_n u\right\|_{L^2(\mathcal{M}_n)} + \left\|\mathcal{S}_{n,\alpha}(c - c')\right\|_{L^2(\mathcal{M}_n)}\right) \\
&\leq \left\|a - a'\right\|_{L^2(\mathcal{M})} C_\sigma\left(1 + \left\|b'\right\|_{L^2(\mathcal{M})} C_\mathcal{R}\left\|u\right\|_{L^2(\mathcal{M})} + \left\|c'\right\|_{L^2(\mathcal{M})}\right) \\
&\quad + \left\|a\right\|_{L^2(\mathcal{M})} L_\sigma\left(2\left\|(b - b') * u\right\|_{L^2(\mathcal{M})} + \left\|b - b'\right\| d + \left\|c - c'\right\|_{L^2(\mathcal{M})}\right) \\
&< dC_\sigma\left(2 + C_\mathcal{R}\left\|u\right\|_{L^2(\mathcal{M})}\right) + 5dL_\sigma = d\left(C_\sigma\left(2 + C_\mathcal{R}\left\|u\right\|_{L^2(\mathcal{M})}\right) + 5L_\sigma\right) = \varepsilon,
\end{aligned}
$$

where the first inequality comes from the computation in Remark 4.15. Hence for all $u \in L^2(\mathcal{M}, \mu)$, $\theta \in \Theta$ and $\varepsilon > 0$ there is a $N_\varepsilon \in \mathbb{N}$ and $\delta > 0$ such that for all $\theta' \in \Theta$ with $d_w(\theta, \theta') < \delta$ and $n \geq N_\varepsilon$ it holds that $\left|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta) - \psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta')\right| < \varepsilon$.

Now we show uniform convergence of $\{\psi(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_{n,\alpha}\}_n$ for all $u \in L^2(\mathcal{M}, \mu)$. Fix $u \in L^2(\mathcal{M}, \mu)$ and $\varepsilon > 0$. The map $\psi(u, \cdot)$ is weakly continuous, and (Θ, d_w) is a compact metric space, so in particular it is uniformly continuous, so there is $d_\varepsilon > 0$, such that for $\theta, \theta' \in \Theta$ with $d_w(\theta, \theta') < d_\varepsilon$ it holds that $|\psi(u, \theta) - \psi(u, \theta')| < \varepsilon/3$. The first part of the proof yields that for every $\theta \in \Theta$, there is a $N_{\varepsilon,\theta} \in \mathbb{N}$ and $\delta_\theta > 0$ such that for all $n \geq N_{\varepsilon,\theta}$ it holds that

$$
\left|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta) - \psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta')\right| < \varepsilon/3 \quad \text{for all } \theta' \in \Theta \text{ with } d_w(\theta, \theta') < \delta_\theta.
$$

Since Θ is compact, the open cover $\Theta \subset \bigcup_{\theta \in \Theta} B_{d_w}(\theta, \delta_\theta)$ has a finite subcover, i.e. there are $\{\theta_k\}_{k=1}^K$ such that $\Theta \subset \bigcup_{k=1}^K B_{d_w}(\theta_k, \delta_{\theta_k})$. Let $N_\varepsilon = \max_{k=1,\dots K} N_{\varepsilon,\theta_k}$. Then for all $n \geq N_\varepsilon$ and all $\theta \in \Theta$ it holds that

$$
|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta) - \psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta_k)| < \varepsilon/3, \quad \text{where } k \in \{1, \dots K\} \text{ s.t. } \theta \in B_{d_w}(\theta_k, \delta_{\theta_k}).
$$

Finally the pointwise convergence result Corollary 4.18 yields that there are $N_{\varepsilon,k} \in \mathbb{N}$ such that for all $k = 1, \dots K$ and $n \geq N_{\varepsilon,k}$ we have

$$
|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta_k) - \psi(u, \theta_k)| < \varepsilon/3,
$$

and thus if we set $N_\varepsilon' := \max_{k=1,\dots K} N_{\varepsilon,k}$, then for all $n \geq N_\varepsilon'$ it holds

$$
|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta_k) - \psi(u, \theta_k)| < \varepsilon/3 \quad \text{for all } k = 1, \dots K.
$$

Putting these together gives for every $n \geq \max\{N_\varepsilon, N_\varepsilon'\}$ and for every $\theta \in \Theta$, that there is a θ_k such that $\theta \in B_{d_w}(\theta_k, \delta_{\theta_k})$ and

$$
\begin{aligned}
|\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta) - \psi(u,\theta)| &\leq |\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta) - \psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta_k)| \\
&\quad + |\psi_n(\mathcal{R}_n u, \mathcal{Q}_{n,\alpha}\theta_k) - \psi(u,\theta_k)| + |\psi(u,\theta_k) - \psi(u,\theta)| \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.
\end{aligned}
$$

Since $\varepsilon > 0$ and $u \in L^2(\mathcal{M}, \mu)$ were arbitrary, this yields

$$
\|\psi_n(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_{n,\alpha} - \psi(u,\cdot)\|_\infty \xrightarrow{n\to\infty} 0 \qquad \text{for all } u \in L^2(\mathcal{M}, \mu), \tag{5.5}
$$

i.e. uniform convergence of the neural response as maps over the parameter space for consistent data. $\qquad\square$

Corollary 5.4. *Let $\nu \in \mathcal{M}(\Theta)$. Then it holds*

$$
|f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu}^{[n]}(\mathcal{R}_n u) - f_\nu(u)| \xrightarrow{n\to\infty} 0 \quad \text{for all } u \in L^2(\mathcal{M}, \mu).
$$

Proof. It holds that

$$
\begin{aligned}
\left|f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu}^{[n]}(\mathcal{R}_n u) - f_\nu(u)\right| &= \left|\langle(\mathcal{Q}_{n,\alpha})_{\#}\nu, \psi_n(\mathcal{R}_n u, \cdot)\rangle_{\mathcal{M}(\Theta_n), C(\Theta_n)} - \langle\nu_n, \psi(u,\cdot)\rangle_{\mathcal{M}(\Theta), C(\Theta, w)}\right| \\
&= \left|\langle\nu, \psi_n(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_{n,\alpha}\rangle_{\mathcal{M}(\Theta), C(\Theta)}\langle\nu_n, \psi(u,\cdot)\rangle_{\mathcal{M}(\Theta), C(\Theta, w)}\right| \\
&= \left|\langle\nu, \psi_n(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_{n,\alpha} - \psi(u,\cdot)\rangle_{\mathcal{M}(\Theta), C(\Theta, w)}\right| \\
&\leq \|\nu\|_{TV}\left\|\psi_n(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_{n,\alpha} - \psi(u,\cdot)\rangle_{\mathcal{M}(\Theta), C(\Theta, w)}\right\|_\infty
\end{aligned}
$$

which goes to 0 as $n \to \infty$ as a consequence of (5.5). $\qquad\square$

Corollary 5.5. *Let $\{\nu_n\}_n \subset \mathcal{M}(\Theta)$ be weak-* converging to a $\nu \in \mathcal{M}(\Theta)$. Then for all $u \in L^2(\mathcal{M}, \mu)$ it holds that*

$$
\lim_{n\to\infty} |f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu_n}^{[n]}(\mathcal{R}_n u) - f_\nu(u)| = 0. \tag{5.6}
$$

Proof. We can write

$$
\left|f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu_n}^{[n]}(\mathcal{R}_n u) - f_\nu(u)\right| \leq \left|f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu_n}^{[n]}(\mathcal{R}_n u) - f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu}^{[n]}(\mathcal{R}_n u)\right| + \left|f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu}^{[n]}(\mathcal{R}_n u) - f_\nu(u)\right|.
$$

The last term converges to 0 by Corollary 5.4. The first term we can split up further

$$
\begin{aligned}
&\left|f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu_n}^{[n]}(\mathcal{R}_n u) - f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu}^{[n]}(\mathcal{R}_n u)\right| \\
&\quad = \left|\langle(\mathcal{Q}_{n,\alpha})_{\#}(\nu_n - \nu), \psi_n(\mathcal{R}_n u, \cdot)\rangle_{\mathcal{M}(\Theta_n), C_0(\Theta_n)}\right| \\
&\quad = \left|\langle\nu_n - \nu, \psi_n(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_{n,\alpha}\rangle_{\mathcal{M}(\Theta), C((\Theta, w))}\right| \\
&\quad \leq \left|\langle\nu_n - \nu, \psi_n(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_{n,\alpha} - \psi(u,\cdot)\rangle_{\mathcal{M}(\Theta), C((\Theta, w))}\right| + \left|\langle\nu_n - \nu, \psi(u,\cdot)\rangle_{\mathcal{M}(\Theta), C((\Theta, w))}\right|.
\end{aligned}
$$

Now, both terms of this bound converge to zero, for the first because we have weak-* convergence $\nu_n \rightharpoonup \nu$ in the space of measures and norm convergence in the space $(C((\Theta, w)), \|\cdot\|_\infty)$ (see (5.5)), and for the second it follows immediately from the weak-* convergence of the measures. $\qquad\square$

We are now ready to prove Γ-convergence (see [4, Def. 1.5] for a definition) of the training functionals:

Theorem 5.6 (Discrete-to-Continuum Γ-convergence of ERM functionals). *It holds that $J_{n,\alpha} \xrightarrow{\Gamma} J_\alpha$ on $\mathcal{M}(\Theta)$.*

Proof. Let $\zeta \geq 0$ arbitrary but fixed.

Lower bound inequality: Let $\nu \in \mathcal{M}(\Theta)$ arbitrary but fixed. Let $\{\nu_n\}_n \subset \mathcal{M}(\Theta)$ such that $\nu_n \xrightarrow{*} \nu$ as $n \to \infty$. W.l.o.g. assume that $\nu_n = (\mathcal{Q}_{n,\alpha}^*)_\# (\mathcal{Q}_{n,\alpha})_\# \nu_n$ for every $n \in \mathbb{N}$, because if it does not hold for infinitely many $n \in \mathbb{N}$, then the desired inequality is satisfied in any case, since $\liminf_{n \to \infty} J_{n,\alpha}(\nu_n) = +\infty \geq J_\alpha(\nu)$, and else the sequence can be reindexed appropriately.

Using (4.14), for all $n \in \mathbb{N}$ we observe that

$$
\begin{aligned}
\|\nu_n\|_{TV} &= \sup_{\substack{\phi \in C((\Theta,w)) \\ \|\phi\|_\infty \leq 1}} |\langle \nu_n, \phi \rangle_{\mathcal{M}(\Theta),C((\Theta,w))}| = \sup_{\substack{\phi \in C((\Theta,w)) \\ \|\phi\|_\infty \leq 1}} |\langle (\mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha})_\# \nu_n, \phi \rangle_{\mathcal{M}(\Theta),C((\Theta,w))}| \\
&= \sup_{\substack{\phi \in C((\Theta,w)) \\ \|\phi\|_\infty \leq 1}} |\langle (\mathcal{Q}_{n,\alpha})_\# \nu_n, \phi \circ \mathcal{Q}_{n,\alpha}^* \rangle_{\mathcal{M}(\Theta_n),C(\Theta_n)}| \\
&= \sup_{\substack{\phi \in C((\mathcal{Q}_{n,\alpha}^* \Theta_n, w)) \\ \|\phi\|_\infty \leq 1}} |\langle (\mathcal{Q}_{n,\alpha})_\# \nu_n, \phi \circ \mathcal{Q}_{n,\alpha}^* \rangle_{\mathcal{M}(\Theta_n),C(\Theta_n)}| \\
&= \sup_{\substack{\phi \in C(\Theta_n) \\ \|\phi\|_\infty \leq 1}} |\langle (\mathcal{Q}_{n,\alpha})_\# \nu_n, \phi \rangle_{\mathcal{M}(\Theta_n),C(\Theta_{n,\alpha})}| = \|(\mathcal{Q}_{n,\alpha})_\# \nu_n\|_{TV}.
\end{aligned}
$$

Furthermore from applying Corollary 5.5 we obtain for all $k \in \{1, \ldots, l\}$ that

$$
\left| f^{[n]}_{(\mathcal{Q}_{n,\alpha})_\# \nu_n}(\mathcal{R}_n u_k) - f_\nu(u_k) \right| \xrightarrow{n \to \infty} 0.
$$

So with that and the continuity of ℓ we may conclude

$$
\begin{aligned}
\liminf_{n \to \infty} J_{n,\alpha}(\nu_n) &= \liminf_{n \to \infty} \frac{1}{l} \sum_{k=1}^l \ell(f_{(\mathcal{Q}_{n,\alpha})_\# \nu_n}(\mathcal{R}_n u_k), y_k) + \zeta \|(\mathcal{Q}_{n,\alpha})_\# \nu_n\|_{TV} \\
&= \frac{1}{l} \sum_{k=1}^l \liminf_{n \to \infty} \ell(f_{(\mathcal{Q}_{n,\alpha})_\# \nu_n}(\mathcal{R}_n u_k), y_k) + \zeta \liminf_{n \to \infty} \|\nu_n\|_{TV} \\
&\geq \frac{1}{l} \sum_{k=1}^l \ell(\lim_{n \to \infty} f^{[n]}_{(\mathcal{Q}_{n,\alpha})_\# \nu_n}(\mathcal{R}_n u_k), y_k) + \zeta \|\nu\|_{TV} \\
&= \frac{1}{l} \sum_{k=1}^l \ell(f_\nu(u_k), y_k) + \zeta \|\nu\|_{TV} = J_\alpha(\nu).
\end{aligned}
$$

Existence of a recovery sequence: Let $\nu \in \mathcal{M}(\Theta)$ arbitrary but fixed, and define the recovery sequence $\{\nu_n\}_n \subset \mathcal{M}(\Theta)$ as $\nu_n := (\mathcal{Q}_{n,\alpha}^*)_\# (\mathcal{Q}_{n,\alpha})_\# \nu$. Note that therefore $(\mathcal{Q}_{n,\alpha})_\# \nu_n = (\mathcal{Q}_{n,\alpha})_\# (\mathcal{Q}_{n,\alpha}^*)_\# (\mathcal{Q}_{n,\alpha})_\# \nu = (\mathcal{Q}_{n,\alpha})_\# \nu$. Because of the pointwise convergence

$$
\left\| \mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha} \theta - \theta \right\|_{(L^2)^3} \xrightarrow{n \to \infty} 0
$$

for all $\theta \in \Theta$, as well as the fact that all $\phi \in C((\Theta, w))$ are uniformly bounded by the compactness of (Θ, w), we have that $\phi \circ \mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha}$ is dominated by the constant function $\|\phi\|_\infty$, i.e.

$$
|\phi(\mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha}(\theta))| \leq \sup_{\vartheta \in \Theta} |\phi(\vartheta)| = \|\phi\|_\infty < \infty,
$$

and thus

$$
\lim_{n \to \infty} \langle \nu_n, \phi \rangle = \lim_{n \to \infty} \langle \nu, \phi \circ (\mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha}) \rangle = \lim_{n \to \infty} \int_\Theta \phi(\mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha} \theta) d\nu = \int_\Theta \phi(\theta) d\nu = \langle \nu, \phi \rangle
$$

for all $\phi \in C((\Theta, w))$. Thus $\nu_n \xrightarrow[n \to \infty]{*} \nu$. Since $\mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha}$ is a continuous operator mapping from $L^2(\mathcal{M}, \mu))^3$ to $L^2(\mathcal{M}, \mu)^3$, so for any $\phi \in C((\Theta, w), \mathbb{R})$ also their composition $\phi \circ \mathcal{Q}_{n,\alpha}^* \mathcal{Q}_{n,\alpha}$ is weakly continuous, therefore

$$
|\langle \nu_n, \phi \rangle| = |\langle (\mathcal{Q}_n^* \mathcal{Q}_n)_\# \nu, \phi \rangle| = |\langle \nu, \phi \circ (\mathcal{Q}_n^* \mathcal{Q}_n) \rangle| \leq \sup_{\substack{\psi \in C(\Theta,w) \\ \|\psi\|_\infty \leq 1}} |\langle \nu, \psi \rangle|.
$$

34

Then

$$\limsup_{n\to\infty}\|\nu_n\|_{TV} = \limsup_{n\to\infty}\sup_{\substack{\phi\in C(\Theta,w)\\\|\phi\|_\infty\le1}}|\langle\nu_n,\phi\rangle| = \limsup_{n\to\infty}\sup_{\substack{\phi\in C(\Theta,w)\\\|\phi\|_\infty\le1}}|\langle(\mathcal{Q}_{n,\alpha}^*\mathcal{Q}_{n,\alpha})_{\#}\nu,\phi\rangle|$$

$$= \limsup_{n\to\infty}\sup_{\substack{\phi\in C(\Theta,w)\\\|\phi\|_\infty\le1}}|\langle\nu,\phi\circ(\mathcal{Q}_{n,\alpha}^*\mathcal{Q}_{n,\alpha})\rangle| \le \limsup_{n\to\infty}\sup_{\substack{\phi\in C(\Theta,w)\\\|\phi\|_\infty\le1}}|\langle\nu,\phi\rangle|$$

$$= \limsup_{n\to\infty}\|\nu\|_{TV} = \|\nu\|_{TV}$$

With the continuity of ℓ, the construction of the recovery sequence and Corollary 5.4 we obtain

$$\limsup_{n\to\infty} J_{n,\alpha}(\nu_n) = \limsup_{n\to\infty}\frac{1}{l}\sum_{k=1}^{l}\ell(f^{[n]}_{(\mathcal{Q}_{n,\alpha})_{\#}\nu_n}(\mathcal{R}_nu_k),y_k) + \zeta\left\|(\mathcal{Q}_{n,\alpha})_{\#}\nu_n\right\|_{TV}$$

$$= \limsup_{n\to\infty}\frac{1}{l}\sum_{k=1}^{l}\ell(f^{[n]}_{(\mathcal{Q}_{n,\alpha})_{\#}\nu}(\mathcal{R}_nu_k),y_k) + \zeta\|\nu_n\|_{TV}$$

$$\le \frac{1}{l}\sum_{k=1}^{l}\ell(\lim_{n\to\infty}f^{[n]}_{(\mathcal{Q}_{n,\alpha})_{\#}\nu}(\mathcal{R}_nu_k),y_k) + \zeta\limsup_{n\to\infty}\|\nu_n\|_{TV}$$

$$\le \frac{1}{l}\sum_{k=1}^{l}\ell(f_\nu(u_k),y_k) + \zeta\|\nu\|_{TV} = J_\alpha(\nu). \qquad\square$$

As a consequence of Theorem 5.6 every sequence of minimizers of $J_{n,\alpha}$ converges weak-* up to a subsequence to a minimizer of J_α, and we also have pointwise convergence of the trained graph networks, parametrized by a minimizer of $\tilde{J}_{n,\alpha}$ to a trained manifold network, up to a subsequence, and consistency of the inputs:

Theorem 5.7 (Uniform convergence of networks on compact sets up to subsequences for consistent inputs and consistent parametrization). *For a fixed $M > 0$ let $\tilde{\rho}_n \in B_{\mathcal{M}(\Theta_n)}(0,M)$ where $n \in \mathbb{N}$, such that there is a $\rho \in \mathcal{M}(\Theta)$, such that $(\mathcal{Q}_{n,\alpha}^*)_{\#}\tilde{\rho}_n =: \rho_n \xrightarrow[n\to\infty]{*} \rho$. Then for every compact $K \subset L^2(\mathcal{M},\mu)$ the sequence of maps*

$$\{f_{\tilde{\rho}_n}\circ\mathcal{R}_n - f_\rho\}_n \subset C((K,\|\cdot\|_{L^2}))$$

is relatively compact in the topology induced by the uniform norm.

Proof. Fix $\varepsilon > 0$, $u \in L^2(\mathcal{M},\mu)$, and $\theta \in \Theta$ and let

$$\delta < \frac{\varepsilon}{L_\sigma C_\mathcal{R}M}.$$

Then we observe from the uniform boundedness of the discretization operators $\mathcal{R}_n$ and the Lipschitzianity of σ that for all $u' \in B_{L^2}(u,\delta)$ it holds

$$|\psi_n(\mathcal{R}_nu,\mathcal{Q}_{n,\alpha}\theta) - \psi_n(\mathcal{R}_nu',\mathcal{Q}_{n,\alpha}\theta)| = |\langle\mathcal{S}_{n,\alpha}a,\sigma(\mathcal{S}_nb*_n\mathcal{R}_nu+\mathcal{S}_{n,\alpha}c) - \sigma(\mathcal{S}_nb*_n\mathcal{R}_nu'+\mathcal{S}_{n,\alpha}c)|$$

$$\le \|\mathcal{S}_{n,\alpha}a\|_{L^2}L_\sigma\|\mathcal{S}_nb*_n(\mathcal{R}_nu-\mathcal{R}_nu')\|_{L^2}$$

$$\le \|a\|_{L^2}L_\sigma\|\mathcal{S}_nb\|_{L^2}\|\mathcal{R}_n(u-u')\|_{L^2}$$

$$\le \|a\|_{L^2}L_\sigma C_\mathcal{R}\|b\|_{L^2}\|u-u'\|_{L^2} < L_\sigma C_\mathcal{R}\delta < \varepsilon/M.$$

This implies that for all for all $u' \in B_\delta(L^2(\mathcal{M},\mu),u)$ we have further

$$\left\|(\psi_n(\mathcal{R}_nu,\cdot) - \psi_n(\mathcal{R}_nu',\cdot))\circ\mathcal{Q}_{n,\alpha}\right\|_\infty = \sup_{\substack{\theta\in\Theta,\\\|\theta\|_\Theta\le1}}|\psi_n(\mathcal{R}_nu,\mathcal{Q}_{n,\alpha}\theta) - \psi_n(\mathcal{R}_nu',\mathcal{Q}_{n,\alpha}\theta)| < \frac{\varepsilon}{M},$$

which finally yields equicontinuity of $\{f_{\tilde{\rho}_n} \circ \mathcal{R}_n\}_n$ since for all $u' \in B_{L^2}(u, \delta)$:

$$
\begin{aligned}
|f_{\tilde{\rho}_n}(\mathcal{R}_n(u)) - f_{\tilde{\rho}_n}(\mathcal{R}_n(u'))| &= |\langle \tilde{\rho}_n, \psi_n(\mathcal{R}_n u, \cdot) - \psi_n(\mathcal{R}_n u', \cdot) \rangle_{\mathcal{M}(\Theta_n), C(\Theta_n)}| \\
&= |\langle \tilde{\rho}_n, (\psi_n(\mathcal{R}_n u, \cdot) - \psi_n(\mathcal{R}_n u', \cdot)) \circ \mathcal{Q}_{n,\alpha} \mathcal{Q}_{n,\alpha}^* \rangle_{\mathcal{M}(\Theta_n), C(\Theta_n)}| \\
&= |\langle (\mathcal{Q}_{n,\alpha}^*)_{\#} \tilde{\rho}_n, (\psi_n(\mathcal{R}_n u, \cdot) - \psi_n(\mathcal{R}_n u', \cdot)) \circ \mathcal{Q}_{n,\alpha} \rangle_{\mathcal{M}(\Theta), C(\Theta)}| \\
&\leq \left\| (\mathcal{Q}_{n,\alpha}^*)_{\#} \tilde{\rho}_n \right\|_{TV} \left\| (\psi_n(\mathcal{R}_n u, \cdot) - \psi_n(\mathcal{R}_n u', \cdot)) \circ \mathcal{Q}_{n,\alpha} \right\|_{\infty} < M \frac{\varepsilon}{M} = \varepsilon.
\end{aligned}
$$

Similarly one observes that for all $n \in \mathbb{N}$ and $u \in L^2(\mathcal{M}, \mu)$, it holds that

$$
\|\psi_n(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_{n,\alpha}\|_{\infty} \leq C_\sigma (1 + C_{\mathcal{R}} \|u\|_{L^2}) < \infty,
$$

and hence $\{f_{\tilde{\rho}_n} \circ \mathcal{R}_n\}_n$ is also pointwise bounded, since then

$$
|f_{\tilde{\rho}_n}(\mathcal{R}_n u)| \leq \left\| (\mathcal{Q}_{n,\alpha}^*)_{\#} \tilde{\rho}_n \right\|_{TV} \|\psi_n(\mathcal{R}_n u, \cdot) \circ \mathcal{Q}_{n,\alpha}\|_{\infty} \leq M C_\sigma (1 + C_{\mathbb{R}} \|u\|_{L^2}) < \infty.
$$

With the above, Arzelà-Ascoli [17, Thm. IV.6.7] yields the claim. $\qquad \square$

So, we may conclude that up to subsequences and w.r.t. consistent data, trained graph convolutional neural networks with consistent parametrization converge on the discretization of compact sets of manifold signals uniformly to a trained manifold convolutional neural network.

Corollary 5.8 (Convergence of trained networks). *Let $M > 0$ and $\{\rho_n\}_n \subset \mathcal{M}(\Theta)$ be a sequence consisting of minimizers ρ_n of $J_{n,\alpha}$ with $\|\rho_n\|_{TV} \leq M$ for all $n \in \mathbb{N}$. Then there exists a subsequence $\{\rho_{n_k}\}_k$ that converges weak-* to a $\rho \in \mathcal{M}(\Theta)$ that is a minimizer of J_α, i.e. $\rho_{n_k} \xrightarrow[k\to\infty]{*} \rho$. Furthermore, $\tilde{\rho} := (\mathcal{Q}_{n,\alpha})_{\#} \rho_n$ is a minimizer of $\tilde{J}_{n,\alpha}$ and it holds that*

  i) $f_{\tilde{\rho}_{n_k}}(\mathcal{R}_{n_k} u) \xrightarrow{k\to\infty} f_\rho(u)$ *for all $u \in L^2(\mathcal{M}, \mu)$, and*
  ii) *for every compact $K \subset L^2(\mathcal{M}, \mu)$ the sequence of maps $\{f_{\tilde{\rho}_{n_k}} \circ \mathcal{R}_{n_k} - f_\rho\}_k \subset C((K, \|\cdot\|_{L^2}))$ is relatively compact in the topology induced by the uniform norm.*

Proof. Since $\{\rho_n\}_n$ is bounded in TV-norm, we can extract a subsequence $\{\rho_{n_k}\}_k$ that converges weak-* to some $\rho \in \mathcal{M}(\Theta)$. Hence, as a consequence of Theorem 5.6, ρ is a minimizer of J_α [4, Thm. 1.21]. Clearly this implies that for any $u \in L^2(\mathcal{M}, \mu)$ we have

$$
f_{\rho_{n_k}}(u) = \langle \rho_{n_k}, \psi(u, \cdot) \rangle_{\mathcal{M}(\Theta), C(\Theta, w)} \xrightarrow{k\to\infty} \langle \rho, \psi(u, \cdot) \rangle_{\mathcal{M}(\Theta), C(\Theta, w)} = f_\rho(u).
$$

Since if ρ_n is a minimizer of $J_{n,\alpha}$ then $\tilde{\rho}_n := (\mathcal{Q}_{n,\alpha})_{\#} \rho_n$ is a minimizer of $\tilde{J}_{n,\alpha}$ and we have $(\mathcal{Q}_{n,\alpha}^*)_{\#} (\mathcal{Q}_{n,\alpha})_{\#} \rho_n = \rho_n$ for every $n \in \mathbb{N}$, hence $(\mathcal{Q}_{n,\alpha}^*)_{\#} \tilde{\rho}_n = \rho_n$. Thus Corollary 5.5 yields pointwise convergence $f_{\tilde{\rho}_{n_k}}(\mathcal{R}_{n_k} u) \xrightarrow{k\to\infty} f_\rho(u)$ for all $u \in L^2(\mathcal{M}, \mu)$, and ii) follows from Theorem 5.7. $\qquad \square$

Remark 5.9. If the regularization parameter $\zeta > 0$, then any sequence $\{\rho_n\}_n \subset \mathcal{M}(\Theta)$ of minimizers of $J_{n,\alpha}$ is bounded by some $M > 0$, hence $\{\rho_n\}_n \subset B_{\mathcal{M}(\Theta)}(0, M)$ and Corollary 5.8 applies.

5.1 Existence of minimizers for the continuum functionals

For completeness, we prove in detail the existence of minimizers of the continuum ERM functionals, even in the absence of the convexity of the loss function ℓ that was assumed in Proposition 5.1.

Theorem 5.10. *J_α is weak-* sequentially lower-semicontinuous.*

Proof. We have to show
$$\liminf_{j\to\infty} J_\alpha(\nu_j) \ge J_\alpha(\nu)$$

for any weak-* convergent sequence $\nu_j \xrightarrow{*} \nu$, where $\nu_j, \nu \in \mathcal{M}(\Theta)$ for all $j \in \mathbb{N}$. We observe by the non-negativity of all summands in $J_\alpha(\cdot)$ and the lower semi-continuity of the total variation norm, and the continuity of the loss function, that

$$\liminf_{j\to\infty} J_\alpha(\nu_j) = \liminf_{j\to\infty} \frac{1}{l}\sum_{k=1}^l \ell(f_{\nu_j}(u_k), y_k) + \zeta \|\nu_j\|_{TV}$$

$$\ge \frac{1}{l}\sum_{k=1}^l \liminf_{j\to\infty} \ell(f_{\nu_j}(u_k), y_k) + \liminf_{j\to\infty} \zeta \|\nu_j\|_{TV}$$

$$\ge \frac{1}{l}\sum_{k=1}^l \ell(\liminf_{j\to\infty} f_{\nu_j}(u_k), y_k) + \zeta \|\nu\|_{TV}.$$

Recall, that f_ρ is of the form $f_\rho(u) = \int_\Theta \psi(u,\theta)d\rho(\theta)$, so since $\psi(u,\cdot) \in C(\Theta, w)$ for all $u \in L^2(\mathcal{M}, \mu)$ we can once again use Riesz-Markov and observe

$$\ell(\liminf_{j\to\infty} f_{\nu_j}(u_k), y_k) = \ell(\lim_{j\to\infty} f_{\nu_j}(u_k), y_k) = \ell(f_\nu(u_k), y_k),$$

since then

$$f_{\nu_j}(u_k) = \int_\Theta \psi(u_k, \theta)d\nu_j(\theta) = \langle \nu_j, \psi(u_k, \cdot)\rangle \xrightarrow{j\to\infty} \langle \nu, \psi(u_k, \cdot)\rangle = f_\nu(u_k).$$

This yields sequential lower semi-continuity of J_α w.r.t. the weak-* topology. $\qquad\square$

Theorem 5.11. *For all $\zeta \ge 0$ and any $M > 0$ the restricted functional $J_\alpha|_{B_{\mathcal{M}(\Theta)}(0,M)}$ admits a minimizer.*

Proof. By definition J_α is bounded from below, since both the fidelity term with the loss function and the regularizing term are nonnegative. Hence, so is its restriction to the ball $B_{\mathcal{M}(\Theta)}(0,M)$, so there exists a minimizing sequence $\{\rho_i\}_i \subset B_{\mathcal{M}(\Theta)}(0, M)$ for the (regularized) ERM functional J_α, meaning

$$J_\alpha(\rho_i) \xrightarrow{i\to\infty} \inf\{J_\alpha(\rho) \mid \rho \in B_{\mathcal{M}(\Theta)}(0, M)\}.$$

We want to show that this minimizing sequence admits a convergent subsequence. Banach-Alaoglu yields weak-* compactness of the ball $B_{\mathcal{M}(\Theta)}(0,M)$, and since $\{\rho_i\}_i \subset B_{\mathcal{M}(\Theta)}(0,M)$, we can extract a weak-* converging subsequence $\{\rho_{i_j}\}_j \subset \{\rho_i\}_i$ with limit $\rho_\infty \in B_M(\mathcal{M}(\Theta))$, i.e. $\rho_{i_j} \xrightarrow{*} \rho_\infty$ for $j \to \infty$. In other words

$$\langle \rho_{i_j}, \psi\rangle_{\mathcal{M}(\Theta), C(\Theta)} \xrightarrow{j\to\infty} \langle \rho_\infty, \psi\rangle_{\mathcal{M}(\Theta), C(\Theta)} \quad \text{for all } \psi \in C(\Theta).$$

Note, that this is

$$\int_\Theta \psi(\theta)d\rho_{i_j}(\theta) \xrightarrow{j\to\infty} \int_\Theta \psi(\theta)d\rho_\infty(\theta) \quad \text{for all } \psi \in C(\Theta).$$

Furthermore J_α is lower-semicontinuous by the previous Theorem 5.10, i.e.

$$\inf\{J_\alpha(\rho) \mid \rho \in B_{\mathcal{M}(\Theta)}(0, M)\} = \lim_{j\to\infty} J(\rho_{i_j}) \ge J(\rho_\infty).$$

Hence $\rho_\infty \in B_M(\mathcal{M}(\Theta))$ minimizes the ERM functional on the unit ball of the Radon measures on the parameter space. $\qquad\square$

Theorem 5.12. *If $\zeta > 0$, then J_α admits a minimizer.*

Proof. Since J_α is bounded from below it admits a minimizing sequence $\{\rho_i\}_i \subset \mathcal{M}(\Theta)$ for the (regularized) ERM functional J_α, meaning

$$J_\alpha(\rho_i) \xrightarrow{i \to \infty} \inf\{J_\alpha(\rho) \mid \rho \in \mathcal{M}(\Theta)\}.$$

By construction, J is bounded from below, hence $\{\rho_i\}_i$ being a minimizing sequence implies boundedness of the sequence $\{J(\rho_i)\}_i$. Since J_α is coercive due to the regularizing term, any minimizing sequence $\{\rho_i\}_i$ of J_α is bounded, so there is a $M \geq 0$ such that $\{\rho_i\}_i \subset B_{\mathcal{M}(\Theta)}(0, M)$, which is weak-* compact, so there is a subsequence $\{\rho_{i_j}\}_j$ and a $\rho_\infty \in B_{\mathcal{M}(\Theta)}(0, M)$, such that $\rho_{i_j} \xrightarrow[j \to \infty]{*} \rho_\infty$. Hence the lower-semicontinuity of J_α (Theorem 5.10 yields

$$\inf\{J_\alpha(\rho) \mid \rho \in \mathcal{M}(\Theta)\} = \lim_{j \to \infty} J_\alpha(\rho_{i_j}) \geq J(\rho_\infty),$$

so ρ_∞ is a minimizer of J_α. $\qquad\square$

6 Discussion

6.1 On the assumptions on the manifold

In this paper we imposed an assumption on the asymptotics of the spectral gaps of the manifold, namely at most polynomial decay. Here, we want to illustrate the delicacy of spectral gap behavior with respect to properties with no intuitive geometric description, such as arithmetic considerations for resonances.

The manifold assumption in geometric machine learning methods builds on possibly leveraging two aspects, a lower intrinsic dimension of the data and also possibly symmetries, which can be correlated with higher multiplicities. As we have seen in Section 3, the appearance of such multiplicities makes the approximation of eigenspaces more delicate than that of eigenvalues: Multiplicities of graph Laplacians need not match those of the Laplace–Beltrami operator, a manifold eigenvalue is approximated by clusters of nearby graph eigenvalues, and the associated combined discrete eigenspaces are expected to converge to the corresponding continuum eigenspace. Furthermore, independent of multiplicities, distinct eigenvalues of the manifold can potentially get arbitrarily close, influencing the ability to distinguish clusters of eigenvalues from the rest of the spectrum, which is required to compare graph eigenfunctions corresponding to the direct sum of the eigenspaces associated to the graph eigenvalue cluster approximating a continuum eigenvalue to the continuum eigenspace. Hence, the error between the interpolation of graph eigenfunctions and suitably aligned corresponding continuum eigenfunctions is controlled by the inverse of the gap between this manifold eigenvalue and its distinct neighbors.

We would like to point out that besides the spectrum being discrete and unbounded, even unbounded rapidly growing multiplicities can occur together with arbitrarily small spectral gaps, due to resonances. To this end we would like to give an example of how small changes in the Riemannian metric can have a large impact on the properties of the spectrum.

Products of spheres. Consider the manifold $\mathbb{S}^2 \times a\mathbb{S}^2$, where $a \in \mathbb{R}$. If $\{k_i\}_{i \in \mathbb{N}}$ enumerate the first eigenvalue of each multiplicity block of $\mathbb{S}^2$, then for $k \in \{k_i, \dots k_{i+1} - 1\}$ the k-th eigenvalue of $\mathbb{S}^2$ is given by $\lambda_k = \lambda_{k_i} = (i-1)i$ [12, p. 35], with multiplicity [39, p. 140]

$$m(\lambda_k) = m(\lambda_{k_i}) = \binom{i+1}{i-1} - \binom{i-1}{i-3} = \frac{(i+1)i - (i-1)(i-2)}{2} = 2i - 1,$$

for $i \geq 3$. Furthermore, if we let $\{l_j\}_{j \in \mathbb{N}}$ enumerate the first eigenvalue of each multiplicity block of $a\mathbb{S}^2$, then for $l \in \{l_j, \dots l_{j+1} - 1\}$ the l-th eigenvalue of $a\mathbb{S}^2$ is given by $\lambda_l^{(a)} = \lambda_{l_j}^{(a)} = (j-1)ja^{-2}$ [12, p. 35].

For product manifolds $\mathcal{M}_1 \times \mathcal{M}_2$ it holds that eigenvalues λ of the Laplacian on $\mathcal{M}_1 \times \mathcal{M}_2$ can be written in the form $\lambda = \lambda^{(1)} + \lambda^{(2)}$ where $\lambda^{(o)}$ is an eigenvalue of the Laplacian on $\mathcal{M}_o$, $o = 1, 2$ [12, pp. 26-27], so in particular we see for the multiplicity of λ that $m(\lambda) \geq m(\lambda^{(1)}) m(\lambda^{(2)}) \geq m(\lambda^{(o)})$ for $o = 1, 2$. So $\{\lambda_k + \lambda_l^{(a)}\}_{k,l \in \mathbb{N}}$ is the spectrum of $\mathbb{S}^2 \times a\mathbb{S}^2$ where the multiplicity of $\lambda_k + \lambda_l^{(a)}$, $l \in \mathbb{N}$, $k \in \{k_i, \ldots, k_{i+1} - 1\}$ is larger or equal to $2i - 1$ for $i \geq 3$. On the other hand, we then see that the gap between the two eigenvalues $\lambda_{k_{i+1}} + \lambda_{l_j}^{(a)}$ and $\lambda_{k_i} + \lambda_{l_{j+1}}^{(a)}$, where $i, j \in \mathbb{N}$ can be calculated as

$$|\lambda_{k_{i+1}} + \lambda_{l_j}^{(a)} - \lambda_{k_i} - \lambda_{l_{j+1}}^{(a)}| = |i(i+1) - i(i-1) + a^{-2}j(j-1) - a^{-2}j(j+1)| = |2i - 2ja^{-2}|.$$

If a is such that $a^2 \in \mathbb{Q}$, then there are $p, q \in \mathbb{Z}$ such that $a^2 = p/q$, hence we obtain

$$|2i - 2ja^{-2}| = 2 \left| \frac{pi - qj}{p} \right| = 2 \frac{|pi - qj|}{p},$$

which is either 0 or greater or equal than $2/p$, since $pi, qj \in \mathbb{Z}$. Since furthermore

$$|\lambda_{l_j}^{(a)} - \lambda_{l_{j+1}}^{(a)}| = 2a^{-2}j \xrightarrow{j \to \infty} +\infty, \quad \text{and} \quad |\lambda_{k_i} - \lambda_{k_{i+1}}| = 2i \xrightarrow{i \to \infty} +\infty,$$

the spectral gaps for $\mathbb{S}^2 \times a\mathbb{S}^2$ are uniformly bounded away from 0.

However, in case $a^2 \notin \mathbb{Q}$, Diophantine approximation [11, pp. 1-3] yields that there are infinitely many $i, j \in \mathbb{N}$ such that $0 < |a^{-2} - \frac{i}{j}| < \frac{1}{j^2}$, hence $0 < |2i - 2ja^{-2}| < \frac{2}{j}$, and therefore in particular for every $\varepsilon > 0$ one can find $i, j \in \mathbb{N}$ such that $0 < |2i - 2ja^{-2}| < \varepsilon$, i.e. so that the eigenvalues $\lambda_{k_{i+1}} + \lambda_{l_j}^{(a)}$ and $\lambda_{k_i} + \lambda_{l_{j+1}}^{(a)}$ are distinct, but the gap between them is smaller than ε. But this means, that for every $\varepsilon > 0$, we can find an eigenvalue λ of $\mathbb{S}^2 \times a\mathbb{S}^2$ whose gap to its distinct neighboring eigenvalues is smaller than ε.

Hence, this provides examples of manifolds with rapidly growing unbounded multiplicities, where seemingly small perturbations of the metric can lead to resonances arising from arithmetic relations among the eigenvalues, that drastically change the spectral properties of the manifold, from spectral gaps being uniformly bounded away from 0 to possibly getting arbitrarily small. The smallness of gaps in this case cannot be inferred from general geometric properties such as curvature bounds, or symmetry, but is instead related to arithmetic resonances of the spectrum.

6.2 On the assumption of Sobolev regularity on the parameters

To establish a formal notion of training convergence for shallow GCNNs, we adopt an infinite-width formulation where the parametrization is represented by measures over the parameter space. To understand such measures as elements of the dual of the space of neural response maps, we require these maps to be continuous with respect to a topology that renders their domain, the parameter space, compact, as was explicated in Section 2.3. While this is easily satisfied in the finite-dimensional setting, the situation is more involved in the infinite-dimensional setting. However, we observed that in the infinite-dimensional setting the convolution term always behaves nicely in that sense and since the activation function is assumed to be Lipschitz-continuous, the neural response maps are always continuous in the parameter variables with respect to the norm topology on L^2. Hence the critical parameters are the first and last. A way to achieve the competing requirements of continuity and compactness in the continuum setting for the first and last parameter is to let them be in the unit ball of a space that is compactly embedded in $L^2(\mathcal{M}, \mu)$, and equip the parameter space with the weak topology. Using this strategy, a natural choice of Θ is therefore the product of the balls $B_{H^\alpha} \times B_{L^2} \times B_{H^\alpha}$, each equipped with its respective weak topology, and with $\alpha > 0$.

Naturally, such a choice induces a restriction on the allowed parameters and learned functions already at the discrete level. To give some interpretation of such restrictions, one can consider the parameters $(a_n, b_n, c_n) \in (\mathbb{R}^n)^3$ for a single discrete neuron, and notice that in each of them the i-th component represents the contribution of node x_i to the parameter's associated signal operation. Since b_n acts through a convolution, it's most naturally seen as a spectral filter on the graph signal u, which can heighten or dampen certain frequency components of the input signal u. Then c_n can be seen as a pointwise shift of the values of the filtered signal, emphasizing or suppressing the contributions of the filtered input signal per node. After that, the activation function is applied node-wise and outputs how much every node of the graph is activated using the spectrally filtered and pointwise shifted signal. Finally a_n, corresponding to a linear outer layer, acts as a spatial readout that aggregates how much of every node should contribute to the final scalar output. With this interpretation, we see that the H^α-regularity is imposed on the two parameters a_n and c_n directly involved in pointwise operations. In consequence, the regularity assumption forces these pointwise responses to not be too localized, which would make it more likely that that specific part of the learned response is strongly dependent on the resolution or realization of the random graph. Moreover, we point to the condition $\varepsilon_n \ll h_n$ which is imposed even in a quantitative form in (2.8), and also implies that sharp differences of values on very close vertices cannot be relevant for any operation built using the edge weights of (2.4).

6.3 On the expressivity of continuum GCNNs

Here we would like to point out that even in the continuum setting and without smoothness restrictions in the parameters, a single neuron cannot detect components of arbitrarily high frequency in its inputs. To see this, note that for orthonormal eigenfunctions $\phi_k \in L^2(\mathcal{M}, \mu)$ it holds for all $\theta = (a, b, c) \in L^2(\mathcal{M}, \mu)^3$

$$
\begin{aligned}
|\psi(\phi_k, \theta) - \psi(0, \theta)| = |\psi(\phi_k, \theta) - \langle a, \sigma(c)\rangle| &= |\langle a, \sigma(b * \phi_k + c) - \sigma(c)\rangle| \\
&\leq \|a\|_{L^2} L_\sigma \|b * \phi_k + c - c\|_{L^2} = \|a\|_{L^2} L_\sigma \|b * \phi_k\|_{L^2} \\
&= \|a\|_{L^2} L_\sigma |\langle b, \phi_k\rangle| \xrightarrow{k \to \infty} 0.
\end{aligned}
$$

Let $g : L^2(\mathcal{M}, \mu)^3 \to \mathbb{R}$ be given by $(a, b, c) = \theta \mapsto g(\theta) = L_\sigma \|a\|_{L^2} \|b\|_{L^2}$. Then

$$
|\psi(\phi_k, \theta) - \langle a, \sigma(c)\rangle| \leq L_\sigma \|a\|_{L^2} \|b * \phi_k\|_{L^2} \leq g(\theta).
$$

Thus by the dominated convergence theorem for any subset $\mathcal{V} \subseteq L^2(\mathcal{M}, \mu)^3$ and $\nu \in \mathcal{M}(\mathcal{V})$ s.t. $\int_\mathcal{V} g(\theta) d\nu(\theta) < \infty$, $\int_\mathcal{V} \psi(\phi_k, \theta) d\nu(\theta) < \infty$ for all eigenfunctions $\phi_k, k \in \mathbb{N}$, and $\int_\mathcal{V} \psi(0, \theta) d\nu(\theta) < \infty$ it is

$$
\begin{aligned}
\left| \int_\mathcal{V} \psi(\phi_k, \theta) d\nu(\theta) - \int_\mathcal{V} \langle a, \sigma(c)\rangle d\nu(\theta) \right| &= \left| \int_\mathcal{V} \psi(\phi_k, \theta) - \langle a, \sigma(c)\rangle d\nu(\theta) \right| \\
&\leq \int_\mathcal{V} |\psi(\phi_k, \theta) - \langle a, \sigma(c)\rangle| d\nu(\theta) \xrightarrow{k \to \infty} 0.
\end{aligned}
$$

Thus

$$
f_\nu(\phi_k) \xrightarrow{k \to \infty} \int_\mathcal{V} \langle a, \sigma(c)\rangle d\nu(\theta) = C_\nu.
$$

Since $\psi(u, \cdot)$ and $\psi(\cdot, \theta)$ are both Lipschitz for all $u \in L^2(\mathcal{M}, \mu)$, $\theta \in \Theta$ it holds that

$$
\begin{aligned}
\left| \psi\left(\sum_{k=K}^{\infty} \langle u, \phi_k\rangle \phi_k, \theta \right) - \psi(0, \theta) \right| &\leq L_\sigma \|a\|_{L^2} \left\| \sum_{k=K}^{\infty} \langle u, \phi_k\rangle \langle b, \phi_k\rangle \phi_k \right\|_{L^2} \\
&\leq L_\sigma \|a\|_{L^2} \sup_{k \geq K} |\langle b, \phi_k\rangle| \|u\|_{L^2} \xrightarrow{K \to \infty} 0.
\end{aligned}
$$

So again dominated convergence yields $|f_\nu \left(\sum_{k=K}^{\infty} \langle u, \phi_k\rangle \phi_k \right) - C_\nu| \xrightarrow{K \to \infty} 0$, so for input signals u that consist only of high frequency components, the convolutional network f_ν for parametrization $\nu \in \mathcal{M}(\mathcal{V})$ will output approximately the same value C_ν.

6.4 Parametrization with probability measures instead of signed measures

We have chosen one particular way to parametrize the networks with signed measures in all parameters, which particularly in the case of activation functions such as ReLU introduces redundancies, which gives rise to many different normalization procedures [19], in particular some leading to working with probability measures. That is also the natural setting for working with Wasserstein gradient flows reflecting first-order optimization on the parameters, as detailed in the next subsection. To this end, define the unreguralized continuum ERM functional

$$E(\nu) := \frac{1}{l} \sum_{k=1}^{l} \ell(f_\nu(u_k), y_k) \quad \text{for } \nu \in \mathcal{M}(\Theta),$$

and the unreguralized discrete ERM functional $E_n : \mathcal{M}(\Theta) \to \mathbb{R}^+ \cup \{+\infty\}$ as

$$E_n(\nu) = \begin{cases} \frac{1}{l} \sum_{k=1}^{l} \ell(f_{(\mathcal{Q}_{n,\alpha})_{\#}\nu}(S_n u_k), y_k), & \nu \in \mathcal{M}(\Theta)_n, \\ +\infty, & \text{otherwise.} \end{cases}$$

Note that the subset of Radon probability measures $\mathscr{P}(\Theta) \subset \mathcal{M}(\Theta)$ is weakly-$*$ closed in the space of finite Radon measures. Thus the liminf inequality still holds if we restrict ourselves to Radon probability measures.

Moreover, for every $\nu \in \mathscr{P}(\Theta)$, we can also find a recovery sequence $\{\nu_n\}_n \subset \mathscr{P}(\Theta)$, i.e. s.t. $\nu_n \xrightarrow[n \to \infty]{*} \nu$ and $E(\nu) \geq \limsup_{n \to \infty} E_n(\nu_n)$. Namely, for $\nu \in \mathscr{P}(\Theta)$ and $n \in \mathbb{N}$, we let $\nu_n := (\mathcal{Q}_n^* \mathcal{Q}_n)_{\#}\nu$. Using the same arguments as for showing the existence of a recovery sequence in the proof for Theorem 5.6, we have $\nu_n \xrightarrow[n \to \infty]{*} \nu$ in $\mathcal{M}(\Theta)$ and by definition the pushforward preserves mass and positivity, hence $\nu_n \in \mathscr{P}(\Theta)$ as well.

6.5 Outlook

One direction for future work would be to consider function outputs instead of scalars, which in combination with the continuum limit would provide a rigorous formalization of an architecture for operator learning on manifolds. One such approach, utilizing the existing neural response map ψ_n would be the approach with measure-valued measures introduced in [16]. In particular, it enables a hypernetwork setup of networks which map any given input into another network (parametrized in the same form), which is trained as a whole using observation operators with finite-dimensional range. Convergence of global minimizers of the training problems on measures, as treated here, would require making concrete the meaning of consistent data in that setting. In this context, it would be even more important to understand which kind of nonlinear operators can be approximated with such a procedure. To that end, the use of Sobolev spaces for parameters as done here makes interpretable characterizations that might be related to PDE solution operators more likely.

Another direction would be more realistic settings for training. While the convex formulation over measures is very convenient, it fails to reflect many additional aspects appearing when training with a procedure involving gradient descent on the parameters themselves. On the one hand, the global minimum may not be attained. On the other, overparametrized situations in which there might be many global minimizers are also not accurately reflected since the purely energy-based approach cannot distinguish which such minimizer is more likely to be reached. Similarly to the way in which working with measures allows for both finite and infinite-width instances simultaneously, formulating the training as a Wasserstein gradient flow on a space of parameter probability measures as in [13] makes optimization insights applicable also in our setting. A natural question to ask would be whether the discrete trajectories converge to the continuum ones under similar data consistency assumptions as the ones we have used here.

References

[1] F. Bach. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18(19):1–53, 2017.

[2] M. Belkin and P. Niyogi. Convergence of Laplacian eigenmaps. In *Proceedings of the 20th International Conference on Neural Information Processing Systems (NeurIPS)*, page 129–136, Cambridge, MA, 2006. MIT Press.

[3] J. Bourgain, H. Brezis, and P. Mironescu. Another look at Sobolev spaces. In *Optimal control and partial differential equations*, pages 439–455. IOS, Amsterdam, 2001.

[4] A. Braides. Γ-*convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.

[5] K. Bredies and M. Carioni. Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calc. Var. Partial Differential Equations*, 59(1):Paper No. 14, 2020.

[6] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[7] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. Preprint arXiv:2104.13478 [cs.LG], 2021.

[8] D. Burago, S. Ivanov, and Y. Kurylev. A graph discretization of the Laplace-Beltrami operator. *Journal of Spectral Theory*, 4, 2015.

[9] J. Calder, D. Slepčev, and M. Thorpe. Rates of convergence for Laplacian semi-supervised learning with low labeling rates. *Res. Math. Sci.*, 10(1):Paper No. 10, 42, 2023.

[10] M. Caselli, E. Florit-Simon, and J. Serra. Fractional Sobolev spaces on Riemannian manifolds. *Math. Ann.*, 390(4):6249–6314, 2024.

[11] J. W. S. Cassels. *An introduction to Diophantine approximation*, volume 45 of *Cambridge tracts in mathematics and mathematical physics*. Cambridge University Press, Cambridge, 1957.

[12] I. Chavel. *Eigenvalues in Riemannian Geometry*, volume 115 of *Pure and applied mathematics*. Academic Press, Inc., Orlando, Florida, 1984.

[13] L. Chizat and F. Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NeurIPS)*, pages 3040–3050, 2018.

[14] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 1997.

[15] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002.

[16] S. Dummer, T. J. Heeringa, and J. A. Iglesias. Vector-valued reproducing kernel Banach spaces for neural networks and operators. Preprint arXiv:2509.26371 [mat.FA], 2025.

[17] N. Dunford and J. T. Schwartz. *Linear Operators Part I: General Theory*, volume VII of *Pure and Applied Mathematics*. Interscience Publishers, Inc., New York, 1966.

[18] M. M. Dunlop, D. Slepčev, A. M. Stuart, and M. Thorpe. Large data and zero noise limits of graph-based semi-supervised learning algorithms. *Appl. Comput. Harmon. Anal.*, 49(2):655–697, 2020.

[19] W. E and S. Wojtowytsch. Representation formulas and pointwise properties for barron functions. *Calc. Var. Partial Differential Equations*, 61, 2022.

[20] F. Gama, E. Isufi, G. Leus, and A. Ribeiro. Graphs, convolutions, and neural networks: From graph filters to graph neural networks. *IEEE Signal Process. Mag.*, 37(6):128–138, 2020.

[21] N. García Trillos, M. Gerlach, M. Hein, and D. Slepčev. Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace-Beltrami operator. *Found. Comput. Math.*, 20(4):827–887, 2020.

[22] N. García Trillos, Z. Kaplan, T. Samakhoana, and D. Sanz-Alonso. On the consistency of graph-based Bayesian semi-supervised learning and the scalability of sampling algorithms. *J. Mach. Learn. Res.*, 21:Paper No. 28, 47, 2020.

[23] N. García Trillos and D. Slepčev. Continuum limit of total variation on point clouds. *Archive for Rational Mechanics and Analysis*, 220(1):193–241, 2015.

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

[25] M. Hein, J. Audibert, and U. von Luxburg. From graphs to manifolds - weak and strong pointwise consistency of graph Laplacians. In *Proceedings of the 18th Conference on Learning Theory (COLT)*, pages 470–485. Springer, Berlin Heidelberg, 2005.

[26] M. Hein, J.-Y. Audibert, and U. von Luxburg. Graph Laplacians and their convergence on random neighborhood graphs. *J. Mach. Learn. Res.*, 8:1325–1370, 2007.

[27] L. Hörmander. The spectral function of an elliptic operator. *Acta Math.*, 121:193–218, 1968.

[28] T. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[29] O. Lablée. *Spectral Theory in Riemannian Geometry*. EMS textbooks in mathematics. European Mathematical Society, 2015.

[30] T. Le and S. Jegelka. Limits, approximation and size transferability for GNNs on sparse graphs via graphops. In *Proceedings of the 37th International Conference on Neural Information Processing Systems (NeurIPS)*, Red Hook, NY, 2023. Curran Associates Inc.

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

[32] R. Levie, W. Huang, L. Bucci, M. Bronstein, and G. Kutyniok. Transferability of spectral graph convolutional neural networks. *J. Mach. Learn. Res.*, 22:Paper No. [272], 59, 2021.

[33] R. Levie, F. Monti, X. Bresson, and M. M. Bronstein. CayleyNets: graph convolutional neural networks with complex rational spectral filters. *IEEE Trans. Signal Process.*, 67(1):97–109, 2019.

[34] S. Maskey, G. Kutyniok, and R. Levie. Generalization bounds for message passing networks on mixture of graphons. *SIAM J. Math. Data Sci.*, 7(2):802–825, 2025.

[35] S. Maskey, R. Levie, and G. Kutyniok. Transferability of graph neural networks: An extended graphon approach. *Applied and Computational Harmonic Analysis*, 63:48–83, 2023.

[36] G. Palatucci, O. Savin, and E. Valdinoci. Local and global minimizers for a variational energy involving a fractional norm. *Annali di Matematica Pura ed Applicata*, 192:673 – 718, 2013.

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

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

[39] E. M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*, volume 32 of *Princeton Mathematical Series*. Princeton University Press, 1971.

[40] M. E. Taylor. *Measure Theory and Integration*, volume 76 of *Graduate studies in mathematics*. American Mathematical Society, 2006.

[41] M. E. Taylor. *Partial differential equations I. Basic theory*, volume 115 of *Applied Mathematical Sciences*. Springer, Cham, third edition, 2023.

[42] J. E. van Engelen and H. H. Hoos. A survey on semi-supervised learning. *Mach. Learn.*, 109(2):373–440, 2020.

[43] V. S. Varadarajan. On the convergence of sample probability distributions. *Sankhyā*, 19:23–26, 1958.

[44] U. von Luxburg. A tutorial on spectral clustering. *Stat. Comput.*, 17(4):395–416, 2007.

[45] Z. Wang, J. Cerviño, and A. Ribeiro. Generalization of geometric graph neural networks with Lipschitz loss functions. *IEEE Trans. Signal Process.*, 73:1549–1561, 2025.

[46] Z. Wang, L. Ruiz, and A. Ribeiro. Convolutional neural networks on manifolds: From graphs and back. *2022 56th Asilomar Conference on Signals, Systems, and Computers*, pages 356–360, 2022.

[47] Z. Wang, L. Ruiz, and A. Ribeiro. Geometric graph filters and neural networks: Limit properties and discriminability trade-offs. *IEEE Trans. Sig. Proc.*, 72:2244–2259, 2024.

[48] D. Werner. Extreme points in spaces of operators and vector-valued measures. *Rend. Circ. Mat. Palermo (2)*, Suppl.(5):135–143, 1984.