Supplemental Materials

A. Proof of Proposition 2

Weyl's law in [35] states that if $\mathcal{M}$ is a compact connected oriented Riemannian manifold of dimension d then

$$N(\lambda) \sim \frac{C_d}{(2\pi)^d} Vol(\mathcal{M}) \lambda^{d/2} \text{ with } N(\lambda) := \#\{\lambda_k \leq \lambda\}. \tag{123}$$

Since eigenvalues of the LB operator $\mathcal{L}$ are $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \cdots$ repeated according to its multiplicity, we can have

$$\lambda_k \sim \frac{(2\pi)^2}{(C_d Vol(\mathcal{M}))^{2/d}} k^{2/d}, \tag{124}$$

where C_d denotes the volume of the unit ball of $\mathbb{R}^d$ and $Vol(\mathcal{M})$ is the volume of manifold $\mathcal{M}$. This indicates that λ_k grows with the same order of the magnitude with $\frac{(2\pi)^2}{(C_d Vol(\mathcal{M}))^{2/d}} k^{2/d}$. With this asymptotic equivalence relationship, we can have

$$\lambda_{k+1} - \frac{(2\pi)^2 (k+1)^{2/d}}{(C_d Vol(\mathcal{M}))^{2/d}} = o\left(\frac{(2\pi)^2 (k+1)^{2/d}}{(C_d Vol(\mathcal{M}))^{2/d}} \right), \tag{125}$$

$$\frac{(2\pi)^2}{(C_d Vol(\mathcal{M}))^{2/d}} k^{2/d} - \lambda_k = o(\lambda_k) \tag{126}$$

Therefore, for any constant $C_1 > 0$, we can find some $K_1(C_1) > 0$, which indicates that K_1 depends on C_1, such that for all $k > K_1(C_1)$, we have

$$\lambda_{k+1} - \frac{(2\pi)^2 (k+1)^{2/d}}{(C_d Vol(\mathcal{M}))^{2/d}} < \frac{C_1 (2\pi)^2 (k+1)^{2/d}}{(C_d Vol(\mathcal{M}))^{2/d}}. \tag{127}$$

Similarly, for any constant $C_2 > 0$, we can find some $K_2(C_2) > 0$, such that for all $k > K_2(C_2)$, we have

$$\frac{(2\pi)^2}{(C_d Vol(\mathcal{M}))^{2/d}} k^{2/d} - \lambda_k < C_2 \lambda_k. \tag{128}$$

Therefore from (127) and (128) we can get upper and lower bound for λ_{k+1} and λ_k respectively. If

$$(1 + C_1)(k+1)^{2/d} - \frac{k^{2/d}}{1 + C_2} \leq \frac{\alpha (Vol(\mathcal{M}) C_d)^{2/d}}{4\pi^2}, \tag{129}$$

we can have $\lambda_{k+1} - \lambda_k \leq \alpha$. The left side can be scaled down to

$$(k+1)^{2/d} - k^{2/d} \geq \min\{1 + C_1, \frac{1}{1 + C_2}\} \frac{2}{d} k^{2/d - 1} = \frac{C_0}{d} k^{2/d-1}$$

This implies that

$$k \geq \left(\frac{\alpha d (Vol(\mathcal{M}) C_d)^{2/d}}{C_0 4\pi^2} \right)^{\frac{d}{2-d}}, \tag{130}$$

with $d > 2$, we can claim that for all $k > K_0(C_0) = \max\{K_1(C_1), K_2(C_2)\}$, if k satisfies

$$k \geq \left\lceil \left(\frac{\alpha d}{C_0 4\pi^2} \right)^{d/(2-d)} (C_d \mathrm{Vol}(\mathcal{M}))^{2/(2-d)} \right\rceil,$$

it holds that $\lambda_{k+1} - \lambda_k \leq \alpha$. Proof of Proposition 3 is similar and is also based on (124).

B. Proof of Proposition 5

Considering that the discrete points $\{x_1, x_2, \ldots, x_n\}$ are uniformly sampled from manifold $\mathcal{M}$ with measure μ, the empirical measure associated with $\mathrm{d}\mu$ can be denoted as $p_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, where δ_{x_i} is the Dirac measure supported on x_i. Similar to the inner product defined in the $L^2(\mathcal{M})$ space (4), the inner product on $L^2(\mathbf{G}_n)$ is denoted as

$$\langle u, v \rangle_{L^2(\mathbf{G}_n)} = \int u(x) v(x) \mathrm{d}p_n = \frac{1}{n} \sum_{i=1}^n u(x_i) v(x_i). \tag{131}$$

The norm in $L^2(\mathbf{G}_n)$ is therefore $\|u\|^2_{L^2(\mathbf{G}_n)} = \langle u, u \rangle_{L^2(\mathbf{G}_n)}$, with $u, v \in L^2(\mathcal{M})$. For signals $\mathbf{u}, \mathbf{v} \in L^2(\mathbf{G}_n)$, the inner product is therefore $\langle \mathbf{u}, \mathbf{v} \rangle_{L^2(\mathbf{G}_n)} = \frac{1}{n} \sum_{i=1}^n [\mathbf{u}]_i [\mathbf{v}]_i$. From here we write $\|\cdot\|_{L^2(\mathbf{G}_n)}$ as $\|\cdot\|$ for simplicity.

We first import the existing results from [47] which indicates the spectral convergence of the constructed Laplacian operator based on the graph $\mathbf{G}_n$ to the LB operator of the underlying manifold.

Theorem 6 (Theorem 2.1 [47])**.** Let $X = \{x_1, x_2, ... x_n\}$ be a set of n points sampled i.i.d. from a d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$. Let $\mathbf{G}_n$ be a graph approximation of $\mathcal{M}$ constructed from X with weight values set as (37) with $t_n = n^{-1/(d+2+\alpha)}$ and $\alpha > 0$. Let $\mathbf{L}_n$ be the graph Laplacian of $\mathbf{G}_n$ and $\mathcal{L}$ be the Laplace-Beltrami operator of $\mathcal{M}$. Let λ_i^n be the i-th eigenvalue of $\mathbf{L}_n$ and $\boldsymbol{\phi}_i^n$ be the corresponding normalized eigenfunction. Let λ_i and $\boldsymbol{\phi}_i$ be the corresponding eigenvalue and eigenfunction of $\mathcal{L}$ respectively. Then, it holds that

$$\lim_{n\to\infty} \lambda_i^n = \lambda_i, \quad \lim_{n\to\infty} |\boldsymbol{\phi}_i^n(x_j) - \boldsymbol{\phi}_i(x_j)| = 0, j = 1, 2 \ldots, n \tag{132}$$

where the limits are taken in probability.

With the definitions of neural networks on graph $\mathbf{G}_n$ and manifold $\mathcal{M}$, the output difference can be written as

$$\begin{aligned} \|\boldsymbol{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{P}_n f) - \mathbf{P}_n \boldsymbol{\Phi}(\mathbf{H}, \mathcal{L}, f))\| &= \left\| \sum_{q=1}^{F_L} \mathbf{x}_L^q - \sum_{q=1}^{F_L} \mathbf{P}_n f_L^q \right\| \\ &\leq \sum_{q=1}^{F_L} \left\| \mathbf{x}_L^q - \mathbf{P}_n f_L^q \right\|. \end{aligned} \tag{133}$$

By inserting the definitions, we have

$$\begin{aligned} &\|\mathbf{x}_l^p - \mathbf{P}_n f_l^p\| \\ &= \left\| \sigma\left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathbf{L}_n) \mathbf{x}_{l-1}^q \right) - \mathbf{P}_n \sigma\left(\sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q \right) \right\| \end{aligned} \tag{134}$$

with $\mathbf{x}_0 = \mathbf{P}_n f$ as the input of the first layer. With a normalized Lipschitz nonlinearity, we have

$$\|\mathbf{x}_l^p - \mathbf{P}_n f_l^p\| \leq \left\| \sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathbf{L}_n) \mathbf{x}_{l-1}^q - \mathbf{P}_n \sum_{q=1}^{F_{l-1}} \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q \right\| \tag{135}$$

$$\leq \sum_{q=1}^{F_{l-1}} \left\| \mathbf{h}_l^{pq}(\mathbf{L}_n) \mathbf{x}_{l-1}^q - \mathbf{P}_n \mathbf{h}_l^{pq}(\mathcal{L}) f_{l-1}^q \right\| \tag{136}$$

The difference can be further decomposed as

$$\begin{aligned}
&\|\mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q - \mathbf{P}_n\mathbf{h}_l^{pq}(\mathcal{L})f_{l-1}^q\| \\
&\quad\leq \|\mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q - \mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{P}_nf_{l-1}^q \\
&\qquad + \mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{P}_nf_{l-1}^q - \mathbf{P}_n\mathbf{h}_l^{pq}(\mathcal{L})f_{l-1}^q\| \quad (137)\\
&\quad\leq \left\|\mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{x}_{l-1}^q - \mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{P}_nf_{l-1}^q\right\| \\
&\qquad + \left\|\mathbf{h}_l^{pq}(\mathbf{L}_n)\mathbf{P}_nf_{l-1}^q - \mathbf{P}_n\mathbf{h}_l^{pq}(\mathcal{L})f_{l-1}^q\right\| \quad (138)
\end{aligned}$$

The first term can be bounded as $\|\mathbf{x}_{l-1}^q - \mathbf{P}_nf_{l-1}^q\|$ with the initial condition $\|\mathbf{x}_0 - \mathbf{P}_nf_0\| = 0$. The second term can be denoted as D_{l-1}^n. With the iteration employed, we can have

$$\|\mathbf{\Phi}(\mathbf{H}, \mathbf{L}_n, \mathbf{P}_nf) - \mathbf{P}_n\mathbf{\Phi}(\mathbf{H}, \mathcal{L}, f)\| \leq \sum_{l=0}^{L}\prod_{l'=l}^{L}F_{l'}D_l^n.$$

Therefore, we can focus on the difference term D_l^n, we omit the feature and layer index to work on a general form.

$$\begin{aligned}
&\|\mathbf{h}(\mathbf{L}_n)\mathbf{P}_nf - \mathbf{P}_n\mathbf{h}(\mathcal{L})f\| \\
&= \left\|\sum_{i=1}^{n}\hat{h}(\lambda_i^n)\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\boldsymbol{\phi}_i^n - \sum_{i=1}^{\infty}\hat{h}(\lambda_i)\langle f, \boldsymbol{\phi}_i\rangle_{\mathcal{M}}\mathbf{P}_n\boldsymbol{\phi}_i\right\|
\end{aligned} \quad (139)$$

We decompose the α-FDT filter function as $\hat{h}(\lambda) = h^{(0)}(\lambda) + \sum_{l\in\mathcal{K}_m}h^{(l)}(\lambda)$ as equations (76) and (77) show. With the triangle inequality and $n > N_\alpha = \max_i\{\lambda_i \in [\Lambda_k(\alpha)]_{k\in\mathcal{K}_s}\}$, we start by analyzing the output difference of $h^{(0)}(\lambda)$ as

$$\begin{aligned}
&\left\|\sum_{i=1}^{N_\alpha}h^{(0)}(\lambda_i^n)\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\boldsymbol{\phi}_i^n - \sum_{i=1}^{N_\alpha}h^{(0)}(\lambda_i)\langle f, \boldsymbol{\phi}_i\rangle_{\mathcal{M}}\mathbf{P}_n\boldsymbol{\phi}_i\right\| \\
&\leq \left\|\sum_{i=1}^{N_\alpha}\left(h^{(0)}(\lambda_i^n) - h^{(0)}(\lambda_i)\right)\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\boldsymbol{\phi}_i^n\right\| \\
&+ \left\|\sum_{i=1}^{N_\alpha}h^{(0)}(\lambda_i)\left(\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\boldsymbol{\phi}_i^n - \langle f, \boldsymbol{\phi}_i\rangle_{\mathcal{M}}\mathbf{P}_n\boldsymbol{\phi}_i\right)\right\|.
\end{aligned} \quad (140)$$

The first term in (140) can be bounded by leveraging the A_h-Lipschitz continuity of the frequency response. From the convergence in probability stated in (132), we can claim that for each eigenvalue $\lambda_i \leq \lambda_{N_\alpha}$, for all $\epsilon_i > 0$ and all $\delta_i > 0$, there exists some N_i such that for all $n > N_i$, we have

$$\mathbb{P}(|\lambda_i^n - \lambda_i| \leq \epsilon_i) \geq 1 - \delta_i, \quad (141)$$

Letting $\epsilon_i < \epsilon$ with $\epsilon > 0$, with probability at least $\prod_{i=1}^{M}(1-\delta_i) := 1 - \delta$, the first term is bounded as

$$\begin{aligned}
&\left\|\sum_{i=1}^{N_\alpha}(h^{(0)}(\lambda_i^n) - h^{(0)}(\lambda_i))\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\boldsymbol{\phi}_i^n\right\| \\
&\leq \sum_{i=1}^{N_\alpha}|h^{(0)}(\lambda_i^n) - h^{(0)}(\lambda_i)||\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}|\|\boldsymbol{\phi}_i^n\| \quad (142)\\
&\leq \sum_{i=1}^{N_\alpha}A_h|\lambda_i^n - \lambda_i|\|\mathbf{P}_nf\|\|\boldsymbol{\phi}_i^n\|^2 \leq N_sA_h\epsilon, \quad (143)
\end{aligned}$$

for all $n > \max\{\max_i N_i, N_\alpha\} := N$.

The second term in (140) can be bounded combined with the convergence of eigenfunctions in (145) as

$$\begin{aligned}
&\left\|\sum_{i=1}^{N_\alpha}h^{(0)}(\lambda_i)\left(\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\boldsymbol{\phi}_i^n - \langle f, \boldsymbol{\phi}_i\rangle_{\mathcal{M}}\mathbf{P}_n\boldsymbol{\phi}_i\right)\right\| \\
&\leq \left\|\sum_{i=1}^{N_\alpha}h^{(0)}(\lambda_i)\left(\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\boldsymbol{\phi}_i^n - \langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\mathbf{P}_n\boldsymbol{\phi}_i\right)\right\| \\
&+ \left\|\sum_{i=1}^{N_\alpha}h^{(0)}(\lambda_i)\left(\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\mathbf{P}_n\boldsymbol{\phi}_i - \langle f, \boldsymbol{\phi}_i\rangle_{\mathcal{M}}\mathbf{P}_n\boldsymbol{\phi}_i\right)\right\|
\end{aligned} \quad (144)$$

From the convergence stated in (132), we can claim that for some fixed eigenfunction $\boldsymbol{\phi}_i$, for all $\epsilon_i > 0$ and all $\delta_i > 0$, there exists some N_i such that for all $n > N_i$, we have

$$\mathbb{P}(|\boldsymbol{\phi}_i^n(x_j) - \boldsymbol{\phi}_i(x_j)| \leq \epsilon_i) \geq 1 - \delta_i, \quad \text{for all } x_j \in X. \quad (145)$$

Therefore, letting $\epsilon_i < \epsilon$ with $\epsilon > 0$, with probability at least $\prod_{i=1}^{M}(1-\delta_i) := 1-\delta$, for all $n > \max\{\max_i N_i, N_\alpha\} := N$, the first term in (144) can be bounded as

$$\begin{aligned}
&\left\|\sum_{i=1}^{N_\alpha}h^{(0)}(\lambda_i)\left(\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\boldsymbol{\phi}_i^n - \langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathcal{M}}\mathbf{P}_n\boldsymbol{\phi}_i\right)\right\| \\
&\qquad\leq \sum_{i=1}^{N_\alpha}\|\mathbf{P}_nf\|\|\boldsymbol{\phi}_i^n - \mathbf{P}_n\boldsymbol{\phi}_i\| \leq N_s\epsilon,
\end{aligned} \quad (146)$$

because the frequency response is non-amplifying as stated in Assumption 1. The last equation comes from the definition of norm in $L^2(\mathbf{G}_n)$. The second term in (144) can be written as

$$\begin{aligned}
&\left\|\sum_{i=1}^{N_\alpha}h^{(0)}(\lambda_i^n)(\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n}\mathbf{P}_n\boldsymbol{\phi}_i - \langle f, \boldsymbol{\phi}_i\rangle_{\mathcal{M}}\mathbf{P}_n\boldsymbol{\phi}_i)\right\| \\
&\leq \sum_{i=1}^{N_\alpha}|h^{(0)}(\lambda_i^n)|\,|\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n} - \langle f, \boldsymbol{\phi}_i\rangle_{\mathcal{M}}|\,\|\mathbf{P}_n\boldsymbol{\phi}_i\|. \quad (147)
\end{aligned}$$

Because $\{x_1, x_2, \cdots, x_n\}$ is a set of uniform sampled points from $\mathcal{M}$, based on Theorem 19 in [48] we can claim that there exists some N such that for all $n > N$

$$\mathbb{P}\left(|\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n} - \langle f, \boldsymbol{\phi}_i\rangle_{\mathcal{M}}| \leq \epsilon\right) \geq 1 - \delta, \quad (148)$$

for all $\epsilon > 0$ and $\delta > 0$. Taking into consider the boundedness of frequency response $|h^{(0)}(\lambda)| \leq 1$ and the bounded energy $\|\mathbf{P}_n\boldsymbol{\phi}_i\|$. Therefore, we have for all $\epsilon > 0$ and $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sum_{i=1}^{N_\alpha}h^{(0)}(\lambda_i^n)\left(\langle\mathbf{P}_nf, \boldsymbol{\phi}_i^n\rangle_{\mathbf{G}_n} - \langle f, \boldsymbol{\phi}_i\rangle_{\mathcal{M}}\right)\mathbf{P}_n\boldsymbol{\phi}_i\right\| \leq N_s\epsilon\right) \\
&\qquad \geq 1 - \delta, \quad (149)
\end{aligned}$$

for all $n > N$.

Combining the above results, we can bound the output difference of $h^{(0)}(\lambda)$. Then we need to analyze the output difference of $h^{(l)}(\lambda)$ and bound this as

$$\begin{aligned}
&\left\|\mathbf{P}_n\mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathbf{L}_n)\mathbf{P}_nf\right\| \\
&\leq \left\|(\hat{h}(C_l) + \delta)\mathbf{P}_nf - (\hat{h}(C_l) - \delta)\mathbf{P}_nf\right\| \leq 2\delta\|\mathbf{P}_nf\|,
\end{aligned} \quad (150)$$

where $\mathbf{h}^{(l)}(\mathcal{L})$ and $\mathbf{h}^{(l)}(\mathbf{L}_n)$ are filters with filter function $h^{(l)}(\lambda)$ on the LB operator $\mathcal{L}$ and graph Laplacian $\mathbf{L}_n$ respectively. Combining the filter functions, we can write

$$\begin{aligned}
&\|\mathbf{P}_n\mathbf{h}(\mathcal{L})f - \mathbf{h}(\mathbf{L}_n)\mathbf{P}_n f\| \\
&= \Bigg\|\mathbf{P}_n\mathbf{h}^{(0)}(\mathcal{L})f + \mathbf{P}_n\sum_{l\in\mathcal{K}_m}\mathbf{h}^{(l)}(\mathcal{L})f - \\
&\qquad \mathbf{h}^{(0)}(\mathbf{L}_n)\mathbf{P}_n f - \sum_{l\in\mathcal{K}_m}\mathbf{h}^{(l)}(\mathbf{L}_n)\mathbf{P}f\Bigg\| \qquad (151)\\
&\leq \|\mathbf{P}_n\mathbf{h}^{(0)}(\mathcal{L})f - \mathbf{h}^{(0)}(\mathbf{L}_n)\mathbf{P}_n f\| + \\
&\qquad \sum_{l\in\mathcal{K}_m}\|\mathbf{P}_n\mathbf{h}^{(l)}(\mathcal{L})f - \mathbf{h}^{(l)}(\mathbf{L}_n)\mathbf{P}_n f\|. \qquad (152)
\end{aligned}$$

Above all, we can claim that there exists some N, such that for all $n > N$, for all $\epsilon' > 0$ and $\delta > 0$, we have

$$\mathbb{P}(\|\mathbf{h}(\mathbf{L}_n)\mathbf{P}_n f - \mathbf{P}_n\mathbf{h}(\mathcal{L})f\| \leq \epsilon') \geq 1 - \delta. \qquad (153)$$

With $\lim_{n\to\infty} D_l^n = 0$ in high probability, this concludes the proof.