WNN approximation by GNN

Alejandro Ribeiro, Luana Ruiz, and Luiz Chamon

1 Assumptions and preliminary results

Before showing formally that as the number of nodes increases the difference between the GNN instatiated from a WNN and the WNN decreases, we state clearly our assumptions and we introduce some basic theorems, corollaries and lemmas necessary for our discussion.

1.1 Assumptions

Assumption 1 *The graphon* $\mathbf{W}$ *is* L_1*-Lipschitz, i.e.* $|\mathbf{W}(u_2, v_2) - \mathbf{W}(u_1, v_1)| \leq L_1(|u_2 - u_1| + |v_2 - v_1|)$.

Assumption 2 *The convolutional filters* h *are* L_2*-Lipschitz and non-amplifying, i.e.* $|h(\lambda)| < 1$.

Assumption 3 *The graphon signal* X *is* L_3*-Lipschitz.*

Assumption 4 *The activation functions are normalized Lipschitz, i.e.* $|\rho(x) - \rho(y)| \leq |x - y|$*, and* $\rho(0) = 0$.

1.2 Definitions

Definition 1 *We define the c-band cardinality B_{nc} of G_n as the number of eigenvalues with absolute value larger than c,*

$$B_{nc} = \#\{\lambda_{ni} : |\lambda_{ni}| > c\} \tag{1}$$

Definition 2 *We define the c-eigenvalue margin δ_{nc} of graph G_n as,*

$$\delta_{nc} = \min_{i,i \neq j} \{|\lambda_{ni} - \lambda_j| : |\lambda_{ni} > c|\} \tag{2}$$

1.3 Important Remarks

To prove Theorem 2, we interpret graphon convolutions as generative models for graph convolutions. Given the graphon $\mathbf{W}(u,v) = \sum_{i \in \mathbb{Z} \setminus \{0\}} \lambda_i \varphi_i(u) \varphi_i(v)$ and a graphon convolution $Y = T_{\mathbf{H}} X$ written as

$$(T_{\mathbf{H}} X)(v) = \sum_{i \in \mathbb{Z} \setminus \{0\}} h(\lambda_i) \varphi_i(v) \int_0^1 \varphi_i(u) X(u) du$$

we can generate graph convolutions $\mathbf{y}_n = \mathbf{H}_n(\mathbf{S}_n)\mathbf{x}_n$ by defining $u_i = (i-1)/n$ for $1 \leq i \leq n$ and setting

$$\begin{aligned}
&[\mathbf{S}_n]_{ij} = \mathbf{W}(u_i, u_j) \\
&[\mathbf{x}_n]_i = X(u_i) \\
&\mathbf{H}_n(\mathbf{S}_n)\mathbf{x}_n = \mathbf{V}_n^{\mathsf{H}} h(\mathbf{\Lambda}_n) \mathbf{V}_n^{\mathsf{H}} \mathbf{x}_n
\end{aligned} \tag{3}$$

where $\mathbf{S}_n$ is the GSO of $\mathbf{G}_n$, the deterministic graph obtained from $\mathbf{W}$, $\mathbf{x}_n$ is the deterministic graph signal obtained by evaluating the graphon signal X at points u_i, and $\mathbf{\Lambda}_n$ and $\mathbf{V}_n$ are the eigenvalues and eigenvectors of $\mathbf{S}_n$ respectively. It is also possible to define graphon convolutions induced by graph convolutions. The graph convolution $\mathbf{y}_n = \mathbf{H}_n(\mathbf{S}_n)\mathbf{x}_n$ induces a graphon convolution $Y_n = T_{\mathbf{H}_n} X_n$ obtained by constructing a partition $I_1 \cup \ldots \cup I_n$ of $[0,1]$ with $I_i = [(i-1)/n, i/n]$ and defining

$$\begin{aligned}
&\mathbf{W}_n(u,v) = [\mathbf{S}_n]_{ij} \times \mathbb{I}(u \in I_i) \mathbb{I}(v \in I_j) \\
&X_n(u) = [\mathbf{x}_n]_i \times \mathbb{I}(u \in I_i) \\
&(T_{\mathbf{H}_n} X_n)(v) = \sum_{i \in \mathbb{Z} \setminus \{0\}} h(\lambda_i^n) \varphi_i^n(v) \int_0^1 \varphi_i^n(u) X_n(u) du
\end{aligned} \tag{4}$$

where $\mathbf{W}_n$ is the *graphon induced by* $\mathbf{G}_n$, X_n is the graphon signal induced by the graph signal $\mathbf{x}_n$ and λ_i^n and φ_i^n are the eigenvalues and eigenfunctions of $\mathbf{W}_n$.

1.4 Graphons as limit objects

To characterize the convergence of a graph sequence $\{\mathbf{G}_n\}$, we consider arbitrary unweighted and undirected graphs $\mathbf{F} = (\mathcal{V}', \mathcal{E}')$ that we call "graph motifs". Homomorphisms of $\mathbf{F}$ into $\mathbf{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ are adjacency preserving maps in which $(i,j) \in \mathcal{E}'$ implies $(i,j) \in \mathcal{E}$. There are $|\mathcal{V}|^{|\mathcal{V}'|} = n^{n'}$ maps from $\mathcal{V}'$ to $\mathcal{V}$, but only some of them are homomorphisms. Hence, we can define a density of homomorphisms $t(\mathbf{F}, \mathbf{G})$, which represents the relative frequency with which the motif $\mathbf{F}$ appears in $\mathbf{G}$.

Homomorphisms of graphs into graphons are defined analogously. Denoting $t(\mathbf{F}, \mathbf{W})$ the density of homomorphisms of the graph $\mathbf{F}$ into the graphon $\mathbf{W}$, we then say that a sequence $\{\mathbf{G}_n\}$ converges to the graphon $\mathbf{W}$ if, for all finite, unweighted and undirected graphs $\mathbf{F}$,

$$\lim_{n\to\infty} t(\mathbf{F}, \mathbf{G}_n) = t(\mathbf{F}, \mathbf{W}). \tag{5}$$

It can be shown that every graphon is the limit object of a convergent graph sequence, and every convergent graph sequence converges to a graphon [1, Chapter 11]. Thus, a graphon identifies an entire collection of graphs. Regardless of their size, these graphs can be considered similar in the sense that they belong to the same "graphon family".

A simple example of convergent graph sequence is obtained by evaluating the graphon. In particular, in this paper we are interested in *deterministic graphs* $\mathbf{G}_n$ constructed by assigning regularly spaced points $u_i = (i-1)/n$ to nodes $1 \leq i \leq n$ and weights $\mathbf{W}(u_i, u_j)$ to edges (i,j), i.e.

$$[\mathbf{S}_n]_{ij} = s_{ij} = \mathbf{W}(u_i, u_j) \tag{6}$$

where $\mathbf{S}_n$ is the adjacency matrix of $\mathbf{G}_n$. A sequence $\{\mathbf{G}_n\}$ generated in this fashion satisfies the condition in (5), therefore $\{\mathbf{G}_n\}$ converges to $\mathbf{W}$ [1, Chapter 11].

1.5 WNNs as deterministic generating models for GNNs

Comparing the GNN and WNN maps $\mathbf{\Phi}(\mathcal{H};\mathbf{S};\mathbf{x})$ and $\mathbf{\Phi}(\mathcal{H};\mathbf{W};X)$, we see that they can have the same set of parameters $\mathcal{H}$. On graphs belonging to a graphon family, this means that GNNs can be built as instantiations of the WNN and, therefore, WNNs can be seen as generative models for GNNs. We consider GNNs $\mathbf{\Phi}(\mathcal{H};\mathbf{S}_n;\mathbf{x}_n)$ built from a WNN $\mathbf{\Phi}(\mathcal{H};\mathbf{W};X)$ by defining $u_i = (i-1)/n$ for $1 \leq i \leq n$ and setting

$$\begin{aligned} &[\mathbf{S}_n]_{ij} = \mathbf{W}(u_i, u_j) \quad \text{and} \\ &[\mathbf{x}_n]_i = X(u_i) \end{aligned} \tag{7}$$

where $\mathbf{S}_n$ is the GSO of $\mathbf{G}_n$, the deterministic graph obtained from $\mathbf{W}$, and $\mathbf{x}_n$ is the *deterministic graph signal* obtained by evaluating the graphon signal X at points u_i. Considering GNNs as instantiations of WNNs is interesting because it allows looking at graphs not as fixed GNN hyperparameters, but as parameters that can be tuned. In other words, it allows GNNs to be adapted both by optimizing the weights in $\mathcal{H}$ and by changing the graph $\mathbf{G}_n$. This makes the learning model scalable and adds flexibility in cases where there are uncertainties associated with the graph.

Conversely, we can also define WNNs induced by GNNs. The WNN induced by a GNN $\mathbf{\Phi}(\mathcal{H};\mathbf{S}_n;\mathbf{x}_n)$ is defined as $\mathbf{\Phi}(\mathcal{H};\mathbf{W}_n;X_n)$, and it is obtained by constructing a partition $I_1 \cup \ldots \cup I_n$ of $[0,1]$ with $I_i = [(i-1)/n, i/n]$ to define

$$\begin{aligned} &\mathbf{W}_n(u,v) = [\mathbf{S}_n]_{ij} \times \mathbb{I}(u \in I_i)\mathbb{I}(v \in I_j) \quad \text{and} \\ &X_n(u) = [\mathbf{x}_n]_i \times \mathbb{I}(u \in I_i) \end{aligned} \tag{8}$$

where $\mathbf{W}_n$ is the *graphon induced by* $\mathbf{G}_n$ and X_n is the *graphon signal induced by the graph signal* $\mathbf{x}_n$. This definition is useful because it allows comparing GNNs with WNNs.

1.6 Preliminary results and lemmas

Proposition 1 *Let* $\mathbf{W} : [0,1]^2 \to [0,1]$ *be an* L_1*-Lipschitz graphon, and let* $\mathbf{W}_n$ *be the graphon induced by the deterministic graph* $\mathbf{G}_n$ *obtained from* $\mathbf{W}$ *as in* (6).

The L_2 norm of $\mathbf{W} - \mathbf{W}_n$ satisfies

$$\|\mathbf{W} - \mathbf{W}_n\|_{L_2([0,1]^2)} \leq \sqrt{\|\mathbf{W} - \mathbf{W}_n\|_{L_1([0,1]^2)}} \leq \frac{\sqrt{L_1}}{\sqrt{n}}.$$

Proof: Refer to the section 4.1. ■

Proposition 2 *Let T and T' be two self-adjoint operators on a separable Hilbert space $\mathcal{H}$ whose spectra are partitioned as $\gamma \cup \Gamma$ and $\omega \cup \Omega$ respectively, with $\gamma \cap \Gamma = \emptyset$ and $\omega \cap \Omega = \emptyset$. If there exists $d > 0$ such that $\min_{x \in \gamma, y \in \Omega} |x - y| \geq d$ and $\min_{x \in \omega, y \in \Gamma} |x - y| \geq d$, then*

$$\|E_T(\gamma) - E_{T'}(\omega)\| \leq \frac{\pi}{2} \frac{\|T - T'\|}{d}$$

Proof: See [2]. ■

Proposition 3 *Let $X \in L_2([0,1])$ be an L_3-Lipschitz graphon signal, and let X_n be the graphon signal induced by the deterministic graph signal $\mathbf{x}_n$ obtained from X as in (7) and (3). The L_2 norm of $X - X_n$ satisfies*

$$\|X - X_n\|_{L_2([0,1])} \leq \frac{L_3}{\sqrt{3n}}.$$

Proof: Refer to the section 4.2. ■

Theorem 1 *Consider the graphon convolution given by $Y = T_{\mathbf{H}} X$, where $h(\lambda)$ has low variability for $|\lambda| < c$. For the graph convolution instantiated from $T_{\mathbf{H}}$ as $\mathbf{y}_n = \mathbf{H}_n(\mathbf{S}_n)\mathbf{x}_n$ [cf. (3)], under Assumptions 1 through 3 it holds*

$$\|Y - Y_n\|_{L_2} \leq \sqrt{L_1} \left(L_2 + \frac{\pi \max\{B_{nc}, B_{mc}\}}{\min\{\delta_{nc}, \delta_{mc}\}} \right) n^{-\frac{1}{2}} \|X\|_{L_2} + \frac{L_3}{\sqrt{3}} n^{-\frac{1}{2}} + L_2 c \|X\| \tag{9}$$

where $Y_n = T_{\mathbf{H}_n} X_n$ is the graph convolution induced by $\mathbf{y}_n = \mathbf{H}_n(\mathbf{S}_n)\mathbf{x}_n$ [cf. (4)], $\max\{B_{nc}, B_{mc}\}$ is the cardinality of the set $\mathcal{C} = \{i \mid |\lambda_i^n| \geq c\}$, and $\min\{\delta_{nc}, \delta_{mc}\} = \min_{i \in \mathcal{C}}(|\lambda_i - \lambda_{i+sgn(i)}^n|, |\lambda_{i+sgn(i)} - \lambda_i^n|, |\lambda_1 - \lambda_{-1}^n|, |\lambda_1^n - \lambda_{-1}|)$, with λ_i and λ_i^n denoting the eigenvalues of $\mathbf{W}$ and $\mathbf{W}_n$ respectively. In particular, if $X = X_n$ we have

$$\|Y - Y_n\|_{L_2} \leq \sqrt{L_1} \left(L_2 + \frac{\pi \max\{B_{nc}, B_{mc}\}}{\min\{\delta_{nc}, \delta_{mc}\}} \right) n^{-\frac{1}{2}} \|X\|_{L_2} + L_2 c \|X\|. \tag{10}$$

Proof: Refer to course website, under lecture 10, *Graph-Graphon filter approximation theorem*. ■

2 Approximation of WNNs with GNNs

Theorem 2 (WNN approximation by GNN) *Consider the L-layer WNN given by $Y = \mathbf{\Phi}(\mathcal{H}; \mathbf{W}; \mathbf{X})$, where $F_0 = F_L = 1$ and $F_\ell = F$ for $1 \leq \ell \leq L-1$. Let the graphon convolutions $h(\lambda)$ be such that $h(\lambda)$ has low variability for $|\lambda| < c$. For the GNN instantiated from this WNN as $\mathbf{y}_n = \mathbf{\Phi}(\mathcal{H}; \mathbf{S}_n; \mathbf{x}_n)$ [cf. (7)], under Assumptions 1 through 4 it holds*

$$\begin{aligned}\|Y_n - Y\|_{L_2} \leq & LF^{L-1}\sqrt{L_1}\left(L_2 + \frac{\pi \max\{B_{nc}, B_{mc}\}}{\min\{\delta_{nc}, \delta_{mc}\}}\right) n^{-\frac{1}{2}} \|X\|_{L_2} \\ & + \frac{L_3}{\sqrt{3}} n^{-\frac{1}{2}} + LF^{L-1} L_2 c \|X\|\end{aligned}$$

where $Y_n = \mathbf{\Phi}(\mathcal{H}; \mathbf{W}_n; X_n)$ is the WNN induced by $\mathbf{y}_n = \mathbf{\Phi}(\mathcal{H}; \mathbf{S}_n; \mathbf{x}_n)$ [cf. (8)], $\max\{B_{nc}, B_{mc}\}$ is the cardinality of the set $\mathcal{C} = \{i \mid |\lambda_i^n| \geq c\}$, and $\min\{\delta_{nc}, \delta_{mc}\} = \min_{i \in \mathcal{C}}(|\lambda_i - \lambda_{i+sgn(i)}^n|, |\lambda_{i+sgn(i)} - \lambda_i^n|, |\lambda_1 - \lambda_{-1}^n|, |\lambda_1^n - \lambda_{-1}|)$, with λ_i and λ_i^n denoting the eigenvalues of $\mathbf{W}$ and $\mathbf{W}_n$ respectively.

3 Proof of Approximation of WNNs with GNNs

Proof: To compute a bound for $\|Y - Y_n\|_{L_2}$, we start by writing it in terms of the last layer's features as

$$\|Y - Y_n\|_{L_2}^2 = \sum_{f=1}^{F_L} \left\| X_L^f - X_{n,L}^f \right\|_{L_2}^2 . \tag{11}$$

At layer ℓ of the WNN $\mathbf{\Phi}(\mathcal{H}; \mathbf{W}; X)$, we have

$$X_\ell^f = \rho\left(\sum_{g=1}^{F_{\ell-1}} \mathbf{h}_\ell^{fg} *_{\mathbf{W}} X_{\ell-1}^g\right) = \rho\left(\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g\right)$$

and similarly for $\mathbf{\Phi}(\mathcal{H}; \mathbf{W}_n; X_n)$,

$$X_{n,\ell}^f = \rho\left(\sum_{g=1}^{F_{\ell-1}} \mathbf{h}_{n,\ell}^{fg} *_{\mathbf{W}} X_{n,\ell-1}^g\right) = \rho\left(\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_{n,\ell}^{fg}} X_{n,\ell-1}^g\right).$$

We can therefore write $\|X_\ell^f - X_{n,\ell}^f\|_{L_2}$ as

$$\left\|X_\ell^f - X_{n,\ell}^f\right\|_{L_2} = \left\|\rho\left(\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g\right) - \rho\left(\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_{n,\ell}^{fg}} X_{n,\ell-1}^g\right)\right\|_{L_2}$$

and, since ρ is normalized Lipschitz,

$$\begin{aligned}\left\|X_\ell^f - X_{n,\ell}^f\right\|_{L_2} &\leq \left\|\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g - T_{\mathbf{H}_{n,\ell}^{fg}} X_{n,\ell-1}^g\right\|_{L_2} \\ &\leq \sum_{g=1}^{F_{\ell-1}} \left\|T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g - T_{\mathbf{H}_{n,\ell}^{fg}} X_{n,\ell-1}^g\right\|_{L_2}.\end{aligned}$$

where the second inequality follows from the triangle inequality. Looking at each feature g independently, we apply the triangle inequality once again to get

$$\begin{aligned}&\left\|T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g - T_{\mathbf{H}_{n,\ell}^{fg}} X_{n,\ell-1}^g\right\|_{L_2} \\ &\leq \left\|T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g - T_{\mathbf{H}_{n,\ell}^{fg}} X_{\ell-1}^g\right\|_{L_2} + \left\|T_{\mathbf{H}_{n,\ell}^{fg}}\left(X_{\ell-1}^g - X_{n,\ell-1}^g\right)\right\|_{L_2}.\end{aligned}$$

The first term on the RHS of this inequality is bounded by (10) in Theorem 1. The second term can be decomposed by using Cauchy-Schwarz and recalling that $|h(\lambda)| < 1$ for all graphon convolutions in the WNN (Assumption 1). We thus obtain a recursion for $\|X_\ell^f - X_{n,\ell}^f\|_{L_2}$, which is given by

$$\begin{aligned}\left\|X_\ell^f - X_{n,\ell}^f\right\|_{L_2} \leq \sum_{g=1}^{F_{\ell-1}} &\sqrt{L_1}\left(L_2 + \frac{\pi \max\{B_{nc}, B_{mc}\}}{\min\{\delta_{nc}, \delta_{mc}\}}\right) n^{-\frac{1}{2}} \|X_{\ell-1}^g\|_{L_2} \\ &+ \sum_{g=1}^{F_{\ell-1}} L_2 c \|X_{\ell-1}^g\| \\ &+ \sum_{g=1}^{F_{\ell-1}} \left\|X_{\ell-1}^g - X_{n,\ell-1}^g\right\|_{L_2}\end{aligned} \tag{12}$$

and whose first term, $\sum_{g=1}^{F_0} \|X_0^g - X_{n,0}^g\|_{L_2} = \sum_{g=1}^{F_0} \|X^g - X_n^g\|_{L_2}$, is bounded as $\sum_{g=1}^{F_0} \|X_0^g - X_{n,0}^g\|_{L_2} \leq F_0 L_3/\sqrt{3n}$ by Proposition 3.

To solve this recursion, we need to compute the norm $\|X_{\ell-1}^g\|_{L_2}$. Since the nonlinearity ρ is normalized Lipschitz and $\rho(0) = 0$ by Assumption 2, this bound can be written as

$$\left\|X_{\ell-1}^g\right\|_{L_2} \leq \left\|\sum_{g=1}^{F_{\ell-1}} T_{\mathbf{H}_\ell^{fg}} X_{\ell-1}^g\right\|_{L_2}$$

and using the triangle and Cauchy Schwarz inequalities,

$$\left\|X_{\ell-1}^g\right\|_{L_2} \leq \sum_{g=1}^{F_{\ell-1}} \left\|T_{\mathbf{H}_\ell^{fg}}\right\|_{L_2} \left\|X_{\ell-1}^g\right\|_{L_2} \leq \sum_{g=1}^{F_{\ell-1}} \left\|X_{\ell-1}^g\right\|_{L_2}$$

where the second inequality follows from $|h(\lambda)| < 1$. Expanding this expression with initial condition $X_0^g = X^g$ yields

$$\left\|X_{\ell-1}^g\right\|_{L_2} \leq \prod_{\ell'=1}^{\ell-1} F_{\ell'} \sum_{g=1}^{F_0} \|X^g\|_{L_2}\,. \tag{13}$$

and substituting it back in (12) to solve the recursion, we get

$$\begin{aligned}\left\|X_\ell^f - X_{n,\ell}^f\right\|_{L_2} \leq & L\sqrt{L_1}\left(L_2 + \frac{\pi \max\{B_{nc}, B_{mc}\}}{\min\{\delta_{nc}, \delta_{mc}\}}\right) n^{-\frac{1}{2}} \left(\prod_{\ell'=1}^{\ell-1} F_{\ell'}\right) \sum_{g=1}^{F_0} \|X^g\|_{L_2} \\ & + L_2 c \left(\prod_{\ell'=1}^{\ell-1} F_{\ell'}\right) \sum_{g=1}^{F_0} \|X^g\|_{L_2} \\ & + \frac{F_0 L_3}{\sqrt{3}} n^{-\frac{1}{2}}.\end{aligned} \tag{14}$$

To arrive at the result of Theorem 2, we evaluate (14) with $\ell = L$ and substitute it into (11) to obtain

$$
\begin{aligned}
\|Y - Y_n\|_{L_2}^2 = &\sum_{f=1}^{F_L} \left\| X_L^f - X_{n,L}^f \right\|_{L_2}^2 \\
\leq &\sum_{f=1}^{F_L} \left(L\sqrt{L_1} \left(L_2 + \frac{\pi \max\{B_{nc}, B_{mc}\}}{\min\{\delta_{nc}, \delta_{mc}\}} \right) n^{-\frac{1}{2}} \left(\prod_{\ell=1}^{L-1} F_\ell \right) \sum_{g=1}^{F_0} \|X^g\|_{L_2} \right. \\
&\left. + L_2 c \left(\prod_{\ell=1}^{L-1} F_\ell \right) \sum_{g=1}^{F_0} \|X^g\|_{L_2} + \frac{F_0 L_3}{\sqrt{3}} n^{-\frac{1}{2}} \right)^2 .
\end{aligned}
\tag{15}
$$

Finally, since $F_0 = F_L = 1$ and $F_\ell = F$ for $1 \leq \ell \leq L-1$,

$$
\begin{aligned}
\|Y - Y_n\|_{L_2} \leq &L\sqrt{L_1} \left(L_2 + \frac{\pi \max\{B_{nc}, B_{mc}\}}{\min\{\delta_{nc}, \delta_{mc}\}} \right) n^{-\frac{1}{2}} F^{L-1} \|X\|_{L_2} \\
&+ LF^{L-1} L_2 c \|X\| + \frac{L_3}{\sqrt{3}} n^{-\frac{1}{2}} .
\end{aligned}
\tag{16}
$$

■

4 Proof of preliminary results and lemmas

4.1 Proof of Proposition 1

Proof: Partitioning the unit interval as $I_i = [(i-1)/n, i/n]$ for $1 \leq i \leq n$ (the same partition used to obtain $\mathbf{S}_n$, and thus $\mathbf{W}_n$, from $\mathbf{W}$), we can use the graphon's Lipschitz property to derive

$$
\|\mathbf{W} - \mathbf{W}_n\|_{L_1(I_i \times I_j)} \leq L_1 \int_0^{1/n} \int_0^{1/n} |u| du dv + L_1 \int_0^{1/n} \int_0^{1/n} |v| dv du = \frac{L_1}{2n^3} + \frac{L_1}{2n^3} = \frac{L_1}{n^3}.
$$

We can then write

$$
\|\mathbf{W} - \mathbf{W}_n\|_{L_1([0,1]^2)} = \sum_{i,j} \|\mathbf{W} - \mathbf{W}_n\|_{L_1(I_i \times I_j)} \leq n^2 \frac{L_1}{n^3} = \frac{L_1}{n}
$$

which, since $\mathbf{W} - \mathbf{W}_n : [0,1]^2 \to [-1,1]$, implies

$$
\|\mathbf{W} - \mathbf{W}_n\|_{L_2([0,1]^2)} \leq \sqrt{\|\mathbf{W} - \mathbf{W}_n\|_{L_1([0,1]^2)}} \leq \frac{\sqrt{L_1}}{\sqrt{n}}.
$$

■

4.2 Proof of Proposition 3

Proof: Partitioning the unit interval as $I_i = [(i-1)/n, i/n]$ for $1 \leq i \leq n$ (the same partition used to obtain $\mathbf{x}_n$, and thus X_n, from X), we can use the Lipschitz property of X to derive

$$\|X - X_n\|_{L_2(I_i)} \leq \sqrt{L_3^2 \int_0^{1/n} u^2 du} = \sqrt{\frac{L_3^2}{3n^3}} + \frac{L_3}{n\sqrt{3n}}.$$

We can then write

$$\|X - X_n\|_{L_2([0,1])} = \sum_i \|X - X_n\|_{L_2(I_i)} \leq n \frac{L_3}{n\sqrt{3n}} = \frac{L_3}{\sqrt{3n}}.$$

■

References

[1] L. Lovász, *Large networks and graph limits*. American Mathematical Society, 2012, vol. 60.

[2] A. Seelmann, "Notes on the sin 2Θ theorem," *Integral Equations and Operator Theory*, vol. 79, no. 4, pp. 579–597, 2014.