▶ Lecture 18 EN.553.744 Prof. Luana Ruiz

Last time: spectral convergence of convolutions

Yet: spectral domain convergence is not enough; graph & graphon convolutions operate in the node domain.

We know:

1) Graph convolutions converge in spectral domain

2) Given $(G_n, x_n) \to (W, x)$, where (W, x) is c-bandlimited, GFTs $\hat{x}_n$ converge to WFT $\hat{x}$.

(THM) Given $(G_n, x_n) \to (W, x)$ with c-bandlimited (W, x); & $H(G_n) = \sum_{k=0}^{K-1} h_k \left(\frac{A_n}{n}\right)^k$ and

$T_H = \sum_{k=0}^{K-1} h_k T_W^{(k)}$, we have: (★)

$$(G_n, y_n) \longrightarrow (W, y)$$

where y_n & y are the graph & graphon conv. outputs resp.

Pf.: $\hat{y}_n \to \hat{y}$ follows $\hat{x}_n \to \hat{x}$ combined with $\hat{H}(G_n) \to \hat{T}_H(W)$. Since x is bandlimited, so is y (convolution is pointwise in the spectral domain). Therefore, the iGFT converges. $\boxtimes$

Yet: This result is still unsatisfactory. Graphons have infinite-dimensional spectra, and still we require the signal to be bandlimited.

Can we do better?

We can guarantee convergence of graph convolutions for any input signal (not only bandlimited) by making filters Lipschitz.

Lipschitz graphon filters:

Filters T_H with Lipschitz spectral response:

$$\boxed{|h(\lambda_1) - h(\lambda_2)| \leq L|\lambda_1 - \lambda_2| \quad \forall \ \lambda_1, \lambda_2 \in [-1, 1]}$$

↳ polynomials, i.e., graphon convolutions, are Lipschitz in bounded intervals

Here, we will consider general, analytic filters $h(\lambda)$ with Lipschitz constant L
(both for generality & to avoid dependence on nb of taps K)

(THM) Given $(G_n, x_n) \to (W, \mathcal{X})$ & $H(G_n) = \sum_{k=0}^{K-1} h_k \left(\frac{A_n}{n}\right)^k$

and $T_H = \sum_{k=0}^{K-1} h_k T_W^{(k)}$ with $h(\lambda)$ Lipschitz, we have:

$$(G_n, y_n) \longrightarrow (W, \mathcal{Y})$$

where y_n & $\mathcal{Y}$ are the graph & graphon conv. outputs resp.

Pf.: WTS $\|\mathcal{Y}_n - \mathcal{Y}\|_2 \to 0$, where $\mathcal{Y}_n$ is the graphon signal induced by y_n

WLOG, we assume $|h(\lambda)| \leq 1 \quad \forall \lambda \in [-1, 1]$

We have: $\mathcal{Y} = \sum_{j \in \mathbb{Z} \setminus \{0\}} h(\lambda_j) \hat{\mathcal{X}}_j \varphi_j$;

$$\mathcal{Y}_n = \sum_{j \in \mathbb{Z} \setminus \{0\}} h(\lambda_j^n) [\hat{\mathcal{X}}_n]_j \varphi_j^n$$

$$\begin{cases} j \in \mathcal{C} = \{ i \text{ s.t. } |\lambda_i| \geq c \} \\ j \notin \mathcal{C} \text{ , where} \end{cases}$$

$$\boxed{c = \frac{1 - |h_0|}{L\,(2\|x\|\,\varepsilon^{-1} + 1)}}$$

with $\varepsilon > 0$, $h_0 = h(0)$ and L the Lip. constant

Then, we can write (by Δ inequality)

$$\|y - y_n\| \leq \Big\| \sum_{j \in e} h(\lambda_j)\,\hat{x}_j\,\varphi_j - \sum_{j \in e} h(\lambda_j^n)\,[\hat{x}_n]_j\,\varphi_j^n \Big\| \quad (1)$$

$$+ \Big\| \sum_{j \notin e} h(\lambda_j)\,\hat{x}_j\,\varphi_j - \sum_{j \notin e} h(\lambda_j^n)\,[\hat{x}_n]_j\,\varphi_j^n \Big\| \quad (2)$$

Note that (1) is equivalent to filter applications to a BL x, with bandwidth c.

Thus, (1) vanishes from (*)

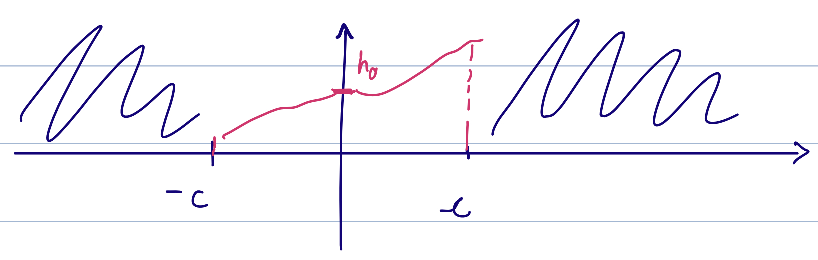
For (2) we do:

$$\left\| \sum_{j \notin e} h(\lambda_j)\, \hat{x}_j\, \varphi_j - \sum_{j \notin e} h(\lambda_j^n)[\hat{x}_n]_j\, \varphi_j^n \right\| \leq$$

$$\left\| \sum_{j \notin e} (h_0 + Lc)\, \hat{x}_j \varphi_j - \sum_{j \notin e} (h_0 - Lc)\, [\hat{x}_n]_j\, \varphi_j^n \right\|$$

$$\leq h_0 \left\| \sum_{j \notin e} \hat{x}_j \varphi_j - [\hat{x}_n]_j\, \varphi_j^n \right\| \quad (3)$$

$$+ \underbrace{Lc \left\| \sum_{j \notin e} \hat{x}_j \varphi_j \right\|}_{Lc\|x\|} + \underbrace{Lc \left\| \sum_{j \notin e} [\hat{x}_n]_j\, \varphi_j^n \right\|}_{= x_n - \sum_{j \in e} [\hat{x}_n]_j \varphi_j^n} \quad (4)$$



$h_{max} = h_0 + Lc$

$h_{min} = h_0 - Lc$

For 3,

$$\Big\|\sum_{j\notin e} \hat{x}_j \varphi_j - [\hat{x}_n]_j \varphi_j^n\Big\| \leqslant \|x - x_n\| + \Big\|\sum_{j\in e} \hat{x}_j \varphi_j - (\hat{x}_n)_j \varphi_j^n\Big\|$$

$$< \varepsilon$$

For 4,

$$\Big\|\sum_{j\notin e} (\hat{x}_n]_j \varphi_j^n\Big\| = \Big\| x_n - \sum_{j\in e} [\hat{x}_n]_j \varphi_j^n \pm x \Big\|$$

$$\leqslant \underbrace{\|x_n - x\| + \Big\|\sum_{j\in e} (\hat{x}_n]_j \varphi_j^n - \hat{x}_j \varphi_j\Big\|}_{\varepsilon} + \underbrace{\Big\|\sum_{j\notin e} \hat{x}_j \varphi_j\Big\|}_{\|x\|}$$

$$< \varepsilon + \|x\|$$

Putting it all together,

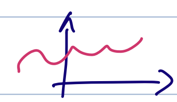
$$2 < h_0 \varepsilon + L_c \|x\| + L_c(\varepsilon + \|x\|) < \varepsilon$$

for all $n > N$. $\blacksquare$

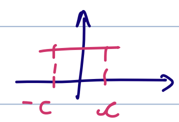
⇒ Graph convolutions converge, but under Lipschitz continuity requirement

↳ As for the GFT, the challenge in showing filter convergence comes eigenvalue accumulation at 0 and its effect on eigenvector convergence

Lipschitz continuity addresses this by ensuring all spectral components near zero are amplified in increasingly similar manner



Recall $c = \dfrac{1-|h_0|}{L\,(2\|x\|\,\varepsilon^{-1}+1)}$;



For fixed c, in order to have $\varepsilon \to 0$ we need progressively smaller L.

At the same time, the smaller c (i.e., the region where spectral components can't be discriminated), the larger we need L to be

↳ convergence-discriminability tradeoff.

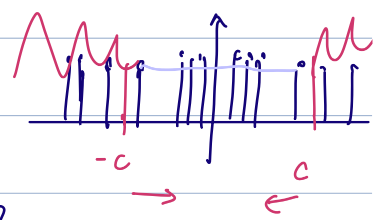
▶ Transferability of graph convolutions & GNNs

Asymptotic convergence reveals important tradeoff, but in order for result to be useful we need finite sample bounds.

We start by introducing two definitions & two assumptions.

(DEFS)

1) c-band cardinality of W:



$$n_c(W) = \# \{ \lambda_i : |\lambda_i| \geq c \}$$

2) c-eigenvalue margin of W, W'

$$\delta_c(W, W') = \min_{i,j \neq i} \{ |\lambda_i(T_{W'}) - \lambda_j(T_W)| : |\lambda_i(T_{W'})| \geq c \}$$

(A S)

3) $\|x\| \leq 1$ (wlog)

4) $|h(\lambda)| < 1$; $h(\lambda)$ L-Lipschitz in $[-1,-c] \cup [c,1]$; l-Lipschitz in $(-c,c)$

(THM) Non-asymptotic convergence of graph convolutions.

Given $(G_n, x_n) \sim (W, x)$ and convolutions

$H(G_n) = \sum_{k=0}^{\infty} h_k \left(\frac{A_n}{n}\right)^k$ and $T_H = \sum_{k=0}^{\infty} h_k T_W^{(k)}$, under

assumptions 3 & 4, we have:

$$\|y_n - y\| \leq \left(L + \frac{\pi n_c}{\delta_c}\right) \|T_W - T_{W_n}\|$$

$$+ (Lc + 2) \|x - x_n\| + 2lc$$

$\Rightarrow$ Convergence $(G_n, x_n) \to (W, x)$ (w/ appropriate node labeling) means approximation improves w/ n as expected

$\Rightarrow$ Convergence - discriminability tradeoff is explicit; larger L & smaller c (= more discriminative filters) lead to higher error bound.

$\Rightarrow$ In the finite-sample regime, unless $\ell = 0$, there is always leftover "nontransferable energy" $2\ell c$ corresponding to spectral components w/ $|\lambda_i| < c$, which do not converge.