Permutation Equivariance of Graph Filters

- We will show that graph convolutional filters are equivariant to permutations

Definition (Permutation matrix)

A square matrix $\mathbf{P}$ is a permutation matrix if it has binary entries so that $\mathbf{P} \in \{0,1\}^{n \times n}$ and it further satisfies $\mathbf{P}\mathbf{1} = \mathbf{1}$ and $\mathbf{P}^T\mathbf{1} = \mathbf{1}$.

- The product $\mathbf{P}^T\mathbf{x}$ reorders the entries of the vector $\mathbf{x}$.
- The product $\mathbf{P}^T\mathbf{S}\mathbf{P}$ is a consistent reordering of the rows and columns of $\mathbf{S}$

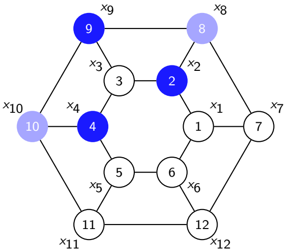
Definition (Permutation matrix)

A square matrix $\mathbf{P}$ is a permutation matrix if it has binary entries so that $\mathbf{P} \in \{0,1\}^{n \times n}$ and it further satisfies $\mathbf{P}\mathbf{1} = \mathbf{1}$ and $\mathbf{P}^T\mathbf{1} = \mathbf{1}$.

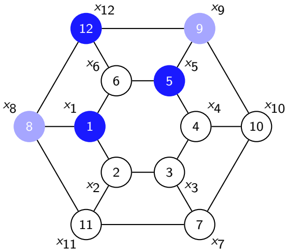
- Since $\mathbf{P}\mathbf{1} = \mathbf{P}^T\mathbf{1} = \mathbf{1}$ with binary entries $\Rightarrow$ Exactly one nonzero entry per row and column of $\mathbf{P}$
- Permutation matrices are unitary $\Rightarrow \mathbf{P}^T\mathbf{P} = \mathbf{I}$. Matrix $\mathbf{P}^T$ undoes the reordering of matrix $\mathbf{P}$

- If $(\mathbf{S}, \mathbf{x})$ is a graph signal, $(\mathbf{P}^T\mathbf{S}\mathbf{P}, \mathbf{P}^T\mathbf{x})$ is a relabeling of $(\mathbf{S}, \mathbf{x})$. Same signal. Different names

Graph signal $\mathbf{x}$ Supported on $\mathbf{S}$



Graph signal $\hat{\mathbf{x}} = \mathbf{P}^T\mathbf{x}$ supported on $\hat{\mathbf{S}} = \mathbf{P}^T\mathbf{S}\mathbf{P}$



- Processing should be label-independent $\Rightarrow$ Permutation equivariance of graph filters and GNNs

- Graph filter $\mathbf{H}(\mathbf{S})$ is a polynomial on shift operator $\mathbf{S}$ with coefficients h_k. Outputs given by

$$\mathbf{H}(\mathbf{S})\mathbf{x} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x}$$

- We consider running the same filter on $(\mathbf{S}, \mathbf{x})$ and permuted (relabeled) $(\hat{\mathbf{S}}, \hat{\mathbf{x}}) = (\mathbf{P}^T\mathbf{S}\mathbf{P}, \mathbf{P}^T\mathbf{x})$

$$\mathbf{H}(\mathbf{S})\mathbf{x} \;=\; \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x} \qquad\qquad \mathbf{H}(\hat{\mathbf{S}})\hat{\mathbf{x}} \;=\; \sum_{k=0}^{K-1} h_k \hat{\mathbf{S}}^k \hat{\mathbf{x}}$$

- Filter $\mathbf{H}(\mathbf{S})\mathbf{x}$ $\Rightarrow$ Coefficients h_k. Input signal $\mathbf{x}$. Instantiated on shift $\mathbf{S}$
- Filter $\mathbf{H}(\hat{\mathbf{S}})\hat{\mathbf{x}}$ $\Rightarrow$ Same Coefficients h_k. Permuted Input signal $\hat{\mathbf{x}}$. Instantiated on permuted shift $\hat{\mathbf{S}}$

Theorem (Permutation equivariance of graph filters)

Consider consistent permutations of the shift operator $\hat{\mathbf{S}} = \mathbf{P}^T\mathbf{S}\mathbf{P}$ and input signal $\hat{\mathbf{x}} = \mathbf{P}^T\mathbf{x}$. Then

$$\mathbf{H}(\hat{\mathbf{S}})\hat{\mathbf{x}} \;=\; \mathbf{P}^T\mathbf{H}(\mathbf{S})\mathbf{x}$$

- Graph filters are equivariant to permutations $\Rightarrow$ Permute input and shift $\equiv$ Permute output

Proof: Write filter output in polynomial form. Use permutation definitions $\hat{\mathbf{S}} = \mathbf{P}^T\mathbf{S}\mathbf{P}$ and $\hat{\mathbf{x}} = \mathbf{P}^T\mathbf{x}$

$$\mathbf{H}(\hat{\mathbf{S}})\hat{\mathbf{x}} = \sum_{k=0}^{K-1} h_k \hat{\mathbf{S}}^k \hat{\mathbf{x}} = \sum_{k=0}^{K-1} h_k \left(\mathbf{P}^T\mathbf{S}\mathbf{P}\right)^k \mathbf{P}^T\mathbf{x}$$

- In the powers $\left(\mathbf{P}^T\mathbf{S}\mathbf{P}\right)^k$, $\mathbf{P}$ and $\mathbf{P}^T$ undo each other $(\mathbf{P}^T\mathbf{P} = \mathbf{I}) \Rightarrow \left(\mathbf{P}^T\mathbf{S}\mathbf{P}\right)^k = \mathbf{P}^T\left(\mathbf{S}\right)^k\mathbf{P}$

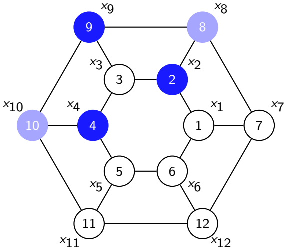
- Substitute this into filter's output expression. Cancel remaining $\mathbf{P}\mathbf{P}^T = \mathbf{I}$ product. Factor $\mathbf{P}^T$

$$\mathbf{H}(\hat{\mathbf{S}})\hat{\mathbf{x}} = \sum_{k=0}^{K-1} h_k \mathbf{P}^T\mathbf{S}^k\mathbf{P}\mathbf{P}^T\mathbf{x} = \sum_{k=0}^{K-1} h_k \mathbf{P}^T\mathbf{S}^k\mathbf{I}\mathbf{x} = \mathbf{P}^T\sum_{k=0}^{K-1} h_k \mathbf{S}^k\mathbf{x} = \mathbf{P}^T\mathbf{H}(\mathbf{S})\mathbf{x} \qquad \blacksquare$$

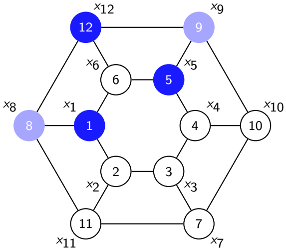
- We requested signal processing independent of labeling $\Rightarrow$ Graph filters fulfill this request

$\Rightarrow$ Permute input and shift $\equiv$ Relabel input $\Rightarrow$ Permute output $\equiv$ Relabel output

Graph signal $\mathbf{x}$ Supported on $\mathbf{S}$



Graph signal $\hat{\mathbf{x}} = \mathbf{P}^T\mathbf{x}$ supported on $\hat{\mathbf{S}} = \mathbf{P}^T\mathbf{S}\mathbf{P}$

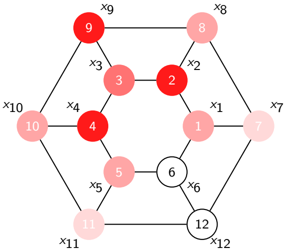


Signal Processing with Graph Filters is Independent of Labeling

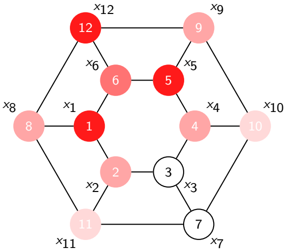
- We requested signal processing independent of labeling $\Rightarrow$ Graph filters fulfill this request

$\Rightarrow$ Permute input and shift $\equiv$ Relabel input $\Rightarrow$ Permute output $\equiv$ Relabel output

Filter's output $\mathbf{H}(\mathbf{S})\mathbf{x}$ Supported on $\mathbf{S}$



Filter's Output $\mathbf{H}(\hat{\mathbf{S}})\hat{\mathbf{x}}$ supported on $\hat{\mathbf{S}}$

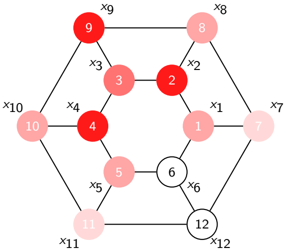


Signal Processing with Graph Filters is Independent of Labeling

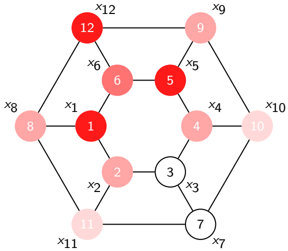
- We requested signal processing independent of labeling $\Rightarrow$ Graph filters fulfill this request

$\Rightarrow$ Permute input and shift $\equiv$ Relabel input $\Rightarrow$ Permute output $\equiv$ Relabel output

Filter's output $\mathbf{H}(\mathbf{S})\mathbf{x}$ Supported on $\mathbf{S}$



Equivariance theorem $\Rightarrow \mathbf{H}(\hat{\mathbf{S}})\hat{\mathbf{x}} = \mathbf{P}^T\mathbf{H}(\mathbf{S})\mathbf{x}$



Permutation Equivariance of Graph Neural Networks

- We will show that graph neural networks inherit the permutation equivariance of graph filters

- L layers recursively process outputs of previous layers. GNN Output parametrized by tensor $\mathcal{H}$

$$\mathbf{x}_\ell = \sigma\left[\sum_{k=0}^{K-1} h_{\ell k}\, \mathbf{S}^k\, \mathbf{x}_{\ell-1}\right] = \sigma\left[\mathbf{H}_\ell(\mathbf{S})\,\mathbf{x}_{\ell-1}\right] \qquad \Phi\Big(\mathbf{x};\ \mathbf{S},\ \mathcal{H}\Big) \ = \ \mathbf{x}_L$$

- We consider running the same GNN on $(\mathbf{S}, \mathbf{x})$ and permuted (relabeled) $(\hat{\mathbf{S}}, \hat{\mathbf{x}}) = (\mathbf{P}^T\mathbf{S}\mathbf{P}, \mathbf{P}^T\mathbf{x})$

$$\Phi\Big(\mathbf{x};\ \mathbf{S},\ \mathcal{H}\Big) \qquad\qquad \Phi\Big(\hat{\mathbf{x}};\ \hat{\mathbf{S}},\ \mathcal{H}\Big)$$

- GNN $\Phi(\mathbf{x}; \mathbf{S}, \mathcal{H}) \ \Rightarrow$ Tensor $\mathcal{H}$. Input signal $\mathbf{x}$. Instantiated on shift $\mathbf{S}$
- GNN $\Phi(\hat{\mathbf{x}}; \hat{\mathbf{S}}, \mathcal{H}) \ \Rightarrow$ Same Tensor $\mathcal{H}$. Permuted Input signal $\hat{\mathbf{x}}$. Instantiated on permuted shift $\hat{\mathbf{S}}$

Theorem (Permutation equivariance of graph neural networks)

Consider consistent permutations of the shift operator $\hat{\mathbf{S}} = \mathbf{P}^T\mathbf{S}\mathbf{P}$ and input signal $\hat{\mathbf{x}} = \mathbf{P}^T\mathbf{x}$. Then

$$\Phi(\hat{\mathbf{x}}; \hat{\mathbf{S}}, \mathcal{H}) \;=\; \mathbf{P}^T\Phi(\mathbf{x}; \mathbf{S}, \mathcal{H})$$

- GNNs equivariant to permutations $\Rightarrow$ Permute input and shift $\equiv$ Permute output

Proof: GNN Layer ℓ recursion on signal $\mathbf{x}_{\ell-1}$ and shift $\mathbf{S}$ $\Rightarrow \mathbf{x}_\ell = \sigma\left[\sum_{k=0}^{K-1} h_{\ell k}\, \mathbf{S}^k\, \mathbf{x}_{\ell-1}\right] = \sigma\left[\mathbf{H}_\ell(\mathbf{S})\mathbf{x}_{\ell-1}\right]$

GNN Layer ℓ recursion on signal $\hat{\mathbf{x}}_{\ell-1}$ and shift $\hat{\mathbf{S}}$ $\Rightarrow \hat{\mathbf{x}}_\ell = \sigma\left[\sum_{k=0}^{K-1} h_{\ell k}\, \hat{\mathbf{S}}^k\, \hat{\mathbf{x}}_{\ell-1}\right] = \sigma\left[\mathbf{H}_\ell(\hat{\mathbf{S}})\hat{\mathbf{x}}_{\ell-1}\right]$

- Assume Layer ℓ inputs satisfy $\hat{\mathbf{x}}_{\ell-1} = \mathbf{P}^T\mathbf{x}_{\ell-1}$. Filters are equivariant. Linearity is pointwise

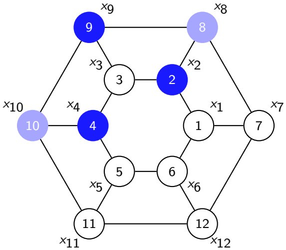
$$\hat{\mathbf{x}}_\ell = \sigma\left[\mathbf{H}_\ell(\hat{\mathbf{S}})\hat{\mathbf{x}}_{\ell-1}\right] = \sigma\left[\mathbf{P}^T\mathbf{H}_\ell(\mathbf{S})\mathbf{x}_{\ell-1}\right] = \mathbf{P}^T\sigma\left[\mathbf{H}_\ell(\mathbf{S})\mathbf{x}_{\ell-1}\right] = \mathbf{P}^T\mathbf{x}_\ell$$

- This in an induction step At Layer 1 we have $\hat{\mathbf{x}} = \mathbf{P}^T\mathbf{x}$ by hypothesis. Induction is complete. ■

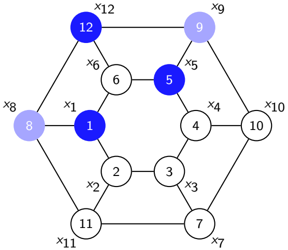
- GNNs, same as graph filters, perform label-independent processing. The nonlinearity is pointwise

$\Rightarrow$ Permute input and shift $\equiv$ Relabel input $\Rightarrow$ Permute output $\equiv$ Relabel output

Graph signal $\mathbf{x}$ Supported on $\mathbf{S}$



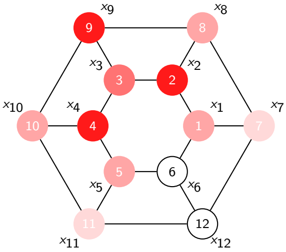
Graph signal $\hat{\mathbf{x}} = \mathbf{P}^T\mathbf{x}$ supported on $\hat{\mathbf{S}} = \mathbf{P}^T\mathbf{S}\mathbf{P}$



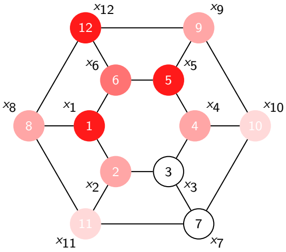
- GNNs, same as graph filters, perform label-independent processing. The nonlinearity is pointwise

$\Rightarrow$ Permute input and shift $\equiv$ Relabel input $\Rightarrow$ Permute output $\equiv$ Relabel output

GNN output $\Phi(\mathbf{x}; \mathbf{S}, \mathcal{H})$ supported on $\mathbf{S}$

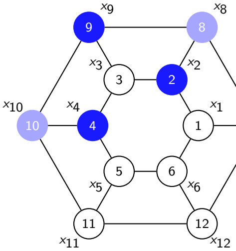


GNN $\Phi(\hat{\mathbf{x}}; \hat{\mathbf{S}}, \mathcal{H}) = \mathbf{P}^T \Phi(\mathbf{x}; \mathbf{S}, \mathcal{H})$ on $\hat{\mathbf{S}} = \mathbf{P}^T \mathbf{S} \mathbf{P}$

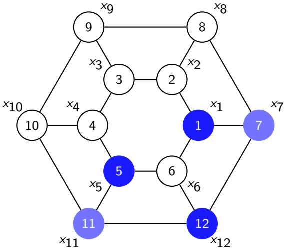


- Equivariance to permutations allows GNNs to exploit symmetries of graphs and graph signals
- By symmetry we mean that the graph can be permuted onto itself $\Rightarrow \mathbf{S} = \mathbf{P}^T\mathbf{S}\mathbf{P}$
- Equivariance theorem implies $\Rightarrow \Phi\Big(\mathbf{P}^T\mathbf{x};\, \mathbf{S}, \mathcal{H}\Big) = \Phi\Big(\mathbf{P}^T\mathbf{x};\, \mathbf{P}^T\mathbf{S}\mathbf{P}, \mathcal{H}\Big) = \mathbf{P}^T\Phi\Big(\mathbf{x};\, \mathbf{S}, \mathcal{H}\Big)$

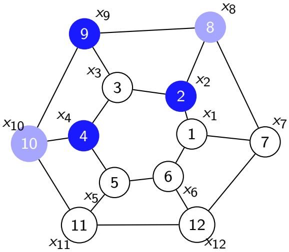
From observing $\mathbf{x}$ supported on $\mathbf{S}$

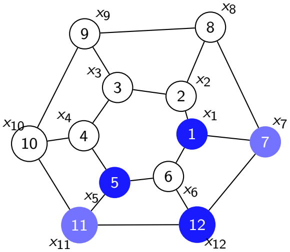


Learn to process $\mathbf{P}^T\mathbf{x}$ supported on $\mathbf{S} = \mathbf{P}^T\mathbf{S}\mathbf{P}$



- Graph not symmetric but close to symmetric $\Rightarrow$ perturbed version of a permutation of itself





- We will show conditions for stability to deformations $\Rightarrow$ Approximate (close to) equivariance

Definition (Operator Distance Modulo Permutation)

For operators Ψ and $\hat{\Psi}$, the operator distance modulo permutation is defined as

$$\left\|\Psi - \hat{\Psi}\right\|_{\mathcal{P}} = \min_{\mathbf{P}\in\mathcal{P}} \max_{\mathbf{x}:\|\mathbf{x}\|=1} \left\| \mathbf{P}^T \Psi(\mathbf{x}) - \hat{\Psi}(\mathbf{P}^T\mathbf{x}) \right\|$$

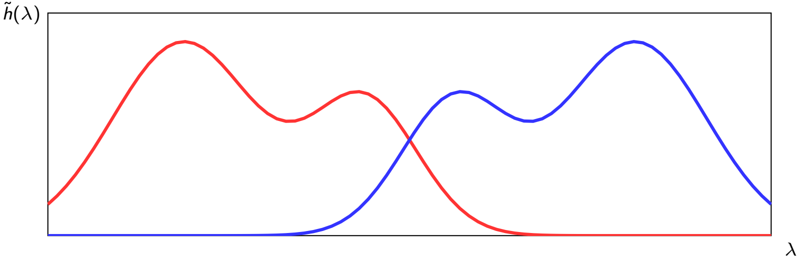
where $\mathcal{P}$ is the set of $n \times n$ permutation matrices and where $\|\cdot\|$ stands for the ℓ_2-norm.

- Equivariance to permutations of graph filters $\Rightarrow$ If $\left\|\hat{\mathbf{S}} - \mathbf{S}\right\|_{\mathcal{P}} = \mathbf{0}$. Then $\left\|\mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S})\right\|_{\mathcal{P}} = \mathbf{0}$
- Equivariance to permutations GNNs $\Rightarrow$ If $\left\|\hat{\mathbf{S}} - \mathbf{S}\right\|_{\mathcal{P}} = \mathbf{0}$. Then $\left\|\Phi(\cdot;\hat{\mathbf{S}},\mathcal{H}) - \Phi(\cdot;\mathbf{S},\mathcal{H})\right\|_{\mathcal{P}} = \mathbf{0}$
- When distance $\left\|\hat{\mathbf{S}} - \mathbf{S}\right\|_{\mathcal{P}}$ is small? (not zero) $\Rightarrow$ Stability properties of graph filters and GNNs

Lipschitz and Integral Lipschitz Filters

- Classes of filters to study discriminablity of GNNs $\Rightarrow$ Lipschitz and integral Lipschitz graph filters

- Graph filters are polynomials on shift operators **S** with given coefficients h_k $\Rightarrow \mathbf{H}(\mathbf{S}) = \sum_{k=0}^{\infty} h_k \mathbf{S}^k$
- Filter's frequency response is the same polynomial with scalar variable λ $\Rightarrow \tilde{h}(\lambda) = \sum_{k=0}^{\infty} h_k \lambda^k$
- Frequency response determined by filter coefficients h_k. Independent of particular given graph



Definition (Lipschitz Filter)

Given a graph filter with coefficients $\mathbf{h} = \{h_k\}_{k=1}^{\infty}$, and graph frequency response

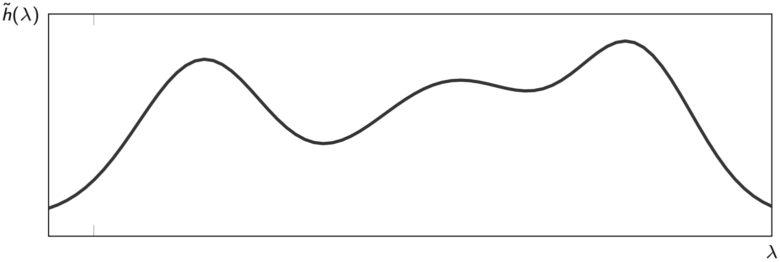
$$\tilde{h}(\lambda) = \sum_{k=0}^{\infty} h_k \lambda^k,$$

we say that the filter is Lipschitz if there exists a constant $C > 0$ such that for λ_1 and λ_2

$$\left| \tilde{h}(\lambda_2) - \tilde{h}(\lambda_1) \right| \leq C \left| \lambda_2 - \lambda_1 \right|.$$

- Change in values of frequency response is at most linear with rate C $\Rightarrow$ Derivative $\tilde{h}'(\lambda) \leq C$

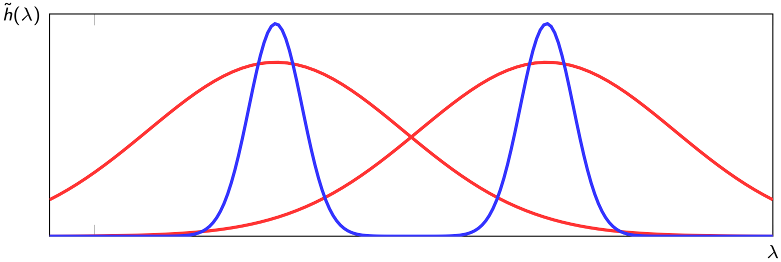
- Frequency response $\tilde{h}(\lambda)$ of Lipschitz filter is Lipschitz continuous $\Rightarrow$ Maximum slope is $\tilde{h}'(\lambda) \leq C$



- Lipschitz constant determines discriminability $\Rightarrow$ Small / Large C $\equiv$ Low / High discriminability

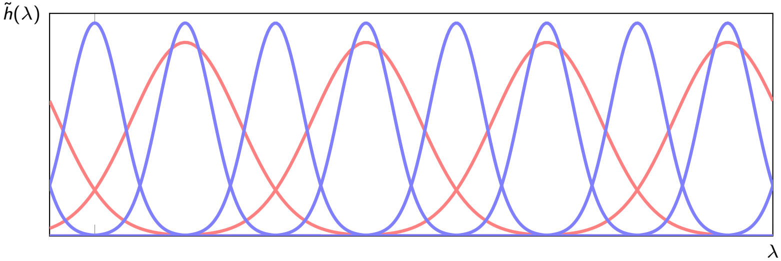
Discriminability of Lipschitz Filters

- Frequency response $\tilde{h}(\lambda)$ of Lipschitz filter is Lipschitz continuous $\Rightarrow$ Maximum slope is $\tilde{h}'(\lambda) \leq C$



- Lipschitz constant determines discriminability $\Rightarrow$ Small / Large C $\equiv$ Low / High discriminability

- A Lipschitz frame with constant C is made up of Lipschitz filters with constant C
- Larger C allows for sharper filters, that can discriminate more signals. Tighter packing
- The discriminability of the frame is (or can be) the same at all frequencies.



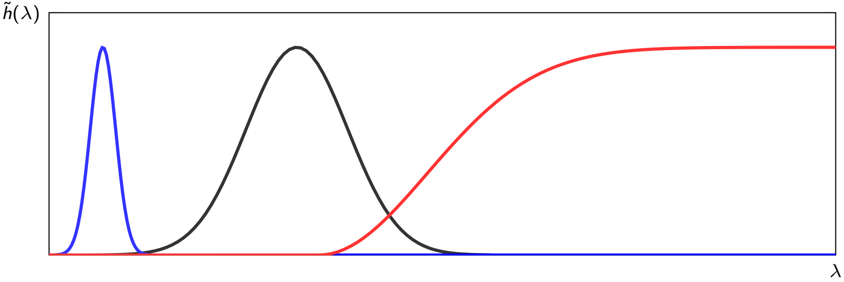
Definition (Integral Lipschitz Filter)

Consider graph filter with coefficients h_k and graph frequency response $\tilde{h}(\lambda) = \sum_{k=0}^{\infty} h_k \lambda^k$. The filter is said integral Lipschitz if there exists constant $C > 0$ such that for all λ_1 and λ_2,

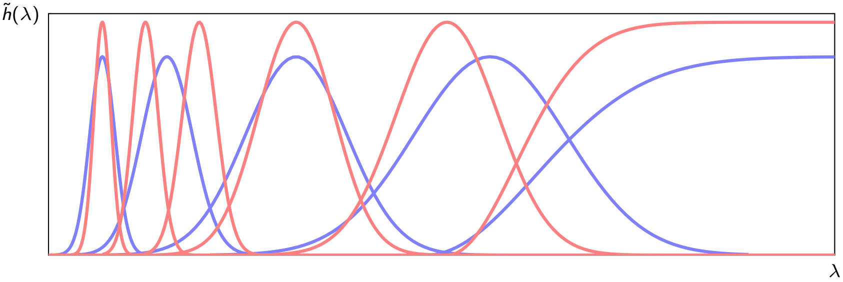
$$|\tilde{h}(\lambda_2) - \tilde{h}(\lambda_1)| \leq C \frac{|\lambda_2 - \lambda_1|}{|\lambda_1 + \lambda_2|/2}.$$

- Lipschitz with a constant that is inversely proportional to the interval's midpoint $\Rightarrow 2C/|\lambda_1 + \lambda_2|$.
- Letting $\lambda_2 \to \lambda_1$ we get that $\lambda \tilde{h}'(\lambda) \leq C$ $\Rightarrow$ The filter can't change for large λ.

- At medium frequencies, integral Lipschitz filters are akin to Lipschitz filters. Roughly speaking
- At low frequencies integral Lipschitz filters can be arbitrarily thin $\Rightarrow$ arbitrary discriminability
- At high frequencies integral Lipschitz filters have to be flat $\Rightarrow$ They lose discriminability



- As Lipschitz frames, integral Lipschitz frames are more discriminative for larger C. Tighter packing
- Except that around $\lambda = 0$, filters can be thin no matter C $\Rightarrow$ High discriminability
- But for large λ filters have to be wide no matter C $\Rightarrow$ No discriminability



Stability of Graph Filters to Scaling

- Scaling of shift operators is a perturbation form that illustrates proof techniques and insights

- We show that graph filters are stable with respect to scaling

Scaling of Shift Operators

- Graphs are subject to estimation error and changes $\Rightarrow$ Running filters on similar graphs

- We scale edges by $(1+\epsilon)$. Scaling deformation of the shift operator $\Rightarrow \hat{\mathbf{S}} = (1+\epsilon)\,\mathbf{S}$

- Deformation model is reasonable $\Rightarrow$ Edges change proportional to their values

- Also unrealistic $\Rightarrow$ All of the edges change by the same proportion

  $\Rightarrow$ Illuminating for discussions. Stability proof contains essential arguments of more generic proof.

Theorem (Integral Lipschitz Graph Filters are Stable to Scaling)

Given graph shift operators $\mathbf{S}$ and $\hat{\mathbf{S}} = (1+\epsilon)\,\mathbf{S}$ and an integral Lipschitz filter with constant C.

The operator norm difference between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded as

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\| \leq C\,\epsilon + \mathcal{O}(\epsilon^2).$$

- Stability to scaling is possible. $\Rightarrow$ But it requires a restriction to the use of integral Lipschitz filters.

- The key arguments of the proof are in the GFT domain. We provide two preliminary spectral facts.

Fact 1:

If $\tilde{\mathbf{x}} = \mathbf{V}^H \mathbf{x}$ is the GFT of $\mathbf{x}$ we can write $\Rightarrow \mathbf{x} = \sum_{i=1}^{n} \tilde{x}_i \mathbf{v}_i$, where $\mathbf{v}_i$ are the eigenvectors of $\mathbf{S}$

Proof: Write $\mathbf{x}$ using the inverse GFT $\Rightarrow$ $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}} = \left[\mathbf{v}_1, \ldots, \mathbf{v}_n\right] \times \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix} = \tilde{x}_1 \mathbf{v}_1 + \ldots + \tilde{x}_n \mathbf{v}_n$

- The key arguments of the proof are in the GFT domain. We provide two preliminary spectral facts.

Fact 2:

The frequency response derivative is $\tilde{h}'(\lambda) = \sum_{k=0}^{\infty} k\, h_k\, \lambda^{k-1}$. Consequently $\lambda\tilde{h}'(\lambda) = \sum_{k=0}^{\infty} k\, h_k\, \lambda^{k}$.

Proof: Frequency response is the series $\Rightarrow \tilde{h}(\lambda) = \sum_{k=0}^{\infty} h_k\, \lambda^k$. The summands' derivatives are $k\, h_k\, \lambda^{k-1}$.

Proof: Filter difference given by graph filter definition $\mathbf{H}(\mathbf{S}) = \sum_{k=0}^{\infty} h_k \mathbf{S}^k$. Further write $\hat{\mathbf{S}} = (1+\epsilon)\,\mathbf{S}$

$$\mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \;=\; \sum_{k=0}^{\infty} h_k \hat{\mathbf{S}}^k - \sum_{k=0}^{\infty} h_k \mathbf{S}^k \;=\; \sum_{k=0}^{\infty} h_k \Big[\big(\,(1+\epsilon)\,\mathbf{S}\,\big)^k - \hat{\mathbf{S}}^k \Big]$$

- Expand binomial $\big(\,(1+\epsilon)\,\mathbf{S}\,\big)^k$ to first order only. Group all high order terms in matrix $\mathbf{O}_k(\epsilon)$

$$\big(\,(1+\epsilon)\,\mathbf{S}\,\big)^k = (1+k\epsilon)\,\mathbf{S}^k + \mathbf{O}_k(\epsilon)$$

- Upon substitution the terms $\mathbf{S}^k$ cancel out $\Rightarrow \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \;=\; \sum_{k=0}^{\infty} h_k k\epsilon \mathbf{S}^k + \mathbf{O}(\epsilon)$

- The matrix $\mathbf{O}(\epsilon)$ satisfies $0 < \lim_{\epsilon \to 0} \dfrac{\|\mathbf{O}(\epsilon)\|}{\epsilon^2} < \infty$ because filter is analytic. Term is of order $\mathcal{O}(\epsilon^2)$

Proof Step 2: Evaluating and Reducing the Operator Norm

- Have reduced the filter difference to $\Rightarrow \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \;=\; \sum_{k=0}^{\infty} h_k k \epsilon \mathbf{S}^k + \mathbf{O}(\epsilon) \;=\; \mathbf{\Delta}(\mathbf{S}) + \mathbf{O}(\epsilon)$

- Where we have defined the filter variation $\mathbf{\Delta}(\mathbf{S}) = \epsilon \sum_{k=0}^{\infty} k h_k \mathbf{S}^k$ to simplify notation

- Triangle inequality $\Rightarrow \|\mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S})\| \;\leq\; \|\mathbf{\Delta}(\mathbf{S})\| + \mathbf{O}(\epsilon) \;=\; \|\mathbf{\Delta}(\mathbf{S})\| + \mathcal{O}(\epsilon^2)$

- Since $\|\mathbf{\Delta}(\mathbf{S})\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{\Delta}(\mathbf{S})\mathbf{x}\|$ theorem follows if we prove $\|\mathbf{\Delta}(\mathbf{S})\mathbf{x}\| \leq C\epsilon$ for all $\mathbf{x}$ with $\|\mathbf{x}\| = 1$

Proof Step 3: Shifting to the GFT Domain

- Product of filter variation with unit norm $\mathbf{x}$. Write the iGFT of the input $\mathbf{x} = \sum_{i=1}^{n} \tilde{x}_i \mathbf{v}_i \quad (\mathbf{S}\mathbf{v}_i = \lambda_i \mathbf{v}_i)$

$$\mathbf{\Delta}(\mathbf{S})\,\mathbf{x} \;=\; \epsilon \sum_{k=0}^{\infty} k\, h_k\, \mathbf{S}^k\, \mathbf{x} \;=\; \epsilon \sum_{k=0}^{\infty} k\, h_k\, \mathbf{S}^k \times \left[\, \sum_{i=1}^{n} \tilde{x}_i \mathbf{v}_i \right] \;=\; \sum_{i=1}^{n} \tilde{x}_i\, \epsilon \sum_{k=0}^{\infty} k\, h_k\, \mathbf{S}^k\, \mathbf{v}_i$$

- Since the $\mathbf{v}_i$ are eigenvectors of $\mathbf{S}$ $\Rightarrow \mathbf{S}^k \mathbf{v}_i = \lambda_i^k \mathbf{v}_i$. With λ_i the associated eigenvalue

$$\mathbf{\Delta}(\mathbf{S})\mathbf{x} \;=\; \epsilon \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} k\, h_k\, \mathbf{S}^k \mathbf{v}_i \;=\; \epsilon \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} k\, h_k\, \lambda_i^k \mathbf{v}_i \;=\; \epsilon \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} k\, h_k\, \lambda_i^k \mathbf{v}_i \;=\; \epsilon \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} k\, h_k\, \lambda_i^k \mathbf{v}_i \;=$$

- The derivative of the filter's response appears $\Rightarrow \sum_{k=0}^{\infty} k\, h_k\, \lambda_i^k \;=\; \lambda_i\, \tilde{h}'(\lambda_i)$

Proof Step 4: Leveraging the Integral Lipschitz Condition

- End up with remarkably simple equation $\Rightarrow \mathbf{\Delta}(\mathbf{S})\mathbf{x} = \epsilon \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} k\, h_k\, \lambda_i^k \mathbf{v}_i = \epsilon \sum_{i=1}^{n} \tilde{x}_i \Big(\lambda_i\, \tilde{h}'(\lambda_i) \Big) \mathbf{v}_i$

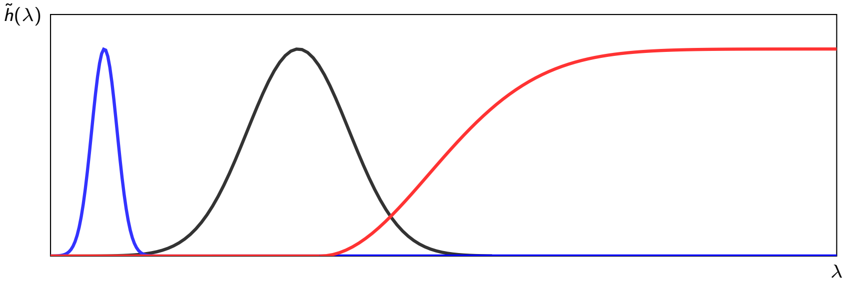
- Which involves the quantity we bound with the integral Lipschitz condition $\Rightarrow \big| \lambda_i\, \tilde{h}'(\lambda_i) \big| \leq C$

- Compute energy. Use integral Lipschitz bound. Recall that signal has unit energy, $\|\mathbf{x}\|^2 = \|\tilde{\mathbf{x}}\|^2 = 1$

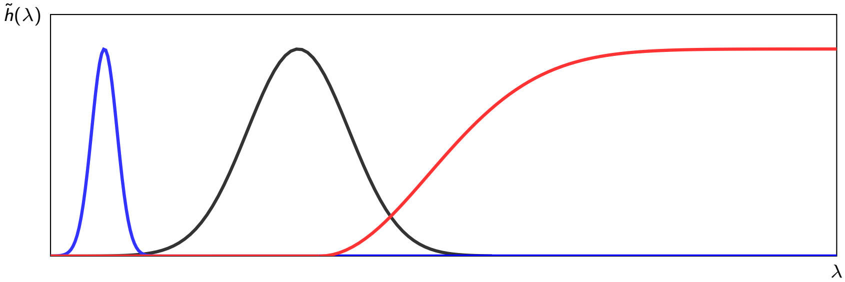
$$\|\mathbf{\Delta}(\mathbf{S})\mathbf{x}\|^2 \;=\; \epsilon^2 \sum_{i=1}^{n} \tilde{x}_i^2 \Big(\lambda_i\, \tilde{h}'(\lambda_i) \Big)^2 \leq \epsilon^2 \sum_{i=1}^{n} \tilde{x}_i^2\, C^2 \;=\; (C\epsilon)^2$$

- Take square root ■

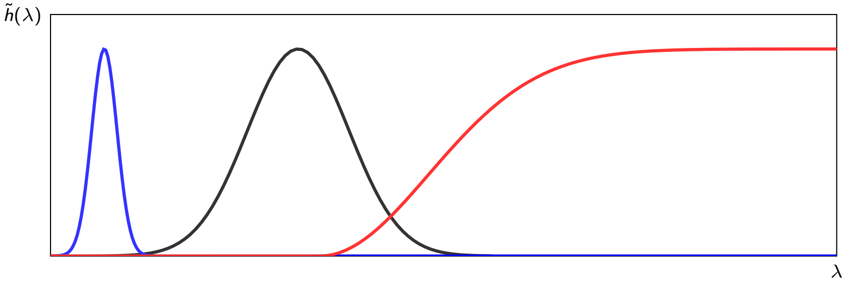
- Integral Lipschitz filters are necessary for stability to deformations of the supporting graph
- This is not an artifact of the analysis. The result is tight. The term $\sum_{k=0}^{\infty} k\, h_k\, \lambda_i^k = \lambda_i\, h'(\lambda_i)$ appears.



- One would expect a stability vs discriminability tradeoff. But in a sense, we get a non-tradeoff.
- Integral Lipschitz filters have to be flat at high frequencies. $\Rightarrow$ They can't discriminate
- It is impossible to separate signals with high frequency features and be stable to deformations



- One would expect a stability vs discriminability tradeoff. But in a sense, we get a non-tradeoff.
- Integral Lipschitz filters have to be flat at high frequencies. $\Rightarrow$ They can't discriminate
- It is impossible to separate signals with high frequency features and be stable to deformations



Stability of Graph Neural Networks to Scaling

- Scaling of shift operators is a perturbation form that illustrates proof techniques and insights
- We show that Graph Neural Networks are stable with respect to scaling

- To avoid appearance of meaningless constants we normalize the filters and the nonlinearity.

- At each layer of the GNN, the filters have unit operator norm $\Rightarrow \big\| \mathbf{H}_\ell(\mathbf{S}) \big\| = 1$

  $\Rightarrow$ Easy to achieve with scaling $\Rightarrow$ Equivalent to $\max_\lambda \tilde{h}_\ell(\lambda) = 1$

- The nonlinearity σ is Lipschitz and normalized so that $\Rightarrow \big\| \sigma(\mathbf{x}_2) - \sigma(\mathbf{x}_1) \big\| \leq \big\| \mathbf{x}_2 - \mathbf{x}_1 \big\|$

  $\Rightarrow$ Easy to achieve with scaling. True of ReLU, hyperbolic tangent, and absolute value

- Joining both assumptions $\Rightarrow$ If input energy is $\|\mathbf{x}\| \leq 1$, all layer outputs have energy $\|\mathbf{x}_\ell\| \leq 1$

Theorem (Integral Lipschitz GNNs are Stable to Scaling)

Given shift operators $\mathbf{S}$ and $\hat{\mathbf{S}} = (1+\epsilon)\,\mathbf{S}$ and a GNN operator $\Phi(\cdot;\mathbf{S},\mathcal{H})$ with L single-feature layers. The filters at each layer have unit operator norms and are integral Lipschitz with constant C. The nonlinearity σ is normalized Lipschitz. Then

$$\Big\|\,\Phi(\cdot;\mathbf{S},\mathcal{H}) - \Phi(\cdot;\hat{\mathbf{S}},\mathcal{H})\,\Big\| \;\leq\; C\,L\,\epsilon + \mathcal{O}(\epsilon^2).$$

- GNNs inherit the stability of graph filters. It's the same bound. Propagated through L layers

Proof: The theorem is true because the nonlinearity is pointwise. It is unaware of the graph.

- Formally $\Rightarrow$ Let $\mathbf{x}_\ell$ be the Layer ℓ output of GNN $\Phi(\mathbf{x}; \mathbf{S}, \mathcal{H})$

  $\Rightarrow$ Let $\hat{\mathbf{x}}_\ell$ be the Layer ℓ output of GNN $\Phi(\hat{\mathbf{x}}; \hat{\mathbf{S}}, \mathcal{H})$

- Layer ℓ is a perceptron with filter $\mathbf{H}_\ell$ $\Rightarrow$ $\left\| \mathbf{x}_\ell - \hat{\mathbf{x}}_\ell \right\| = \left\| \sigma\left[\mathbf{H}_\ell(\mathbf{S})\mathbf{x}_{\ell-1}\right] - \sigma\left[\mathbf{H}_\ell(\hat{\mathbf{S}})\hat{\mathbf{x}}_{\ell-1}\right] \right\|$

- Nonlinearity is normalized Lipschitz $\Rightarrow$ $\left\| \mathbf{x}_\ell - \hat{\mathbf{x}}_\ell \right\| \leq \left\| \mathbf{H}_\ell(\mathbf{S})\mathbf{x}_{\ell-1} - \mathbf{H}_\ell(\hat{\mathbf{S}})\hat{\mathbf{x}}_{\ell-1} \right\|$

- This is the critical step of the proof. The rest of the proof is just algebra.

Proof Step 2: Implementing Norm Manipulations

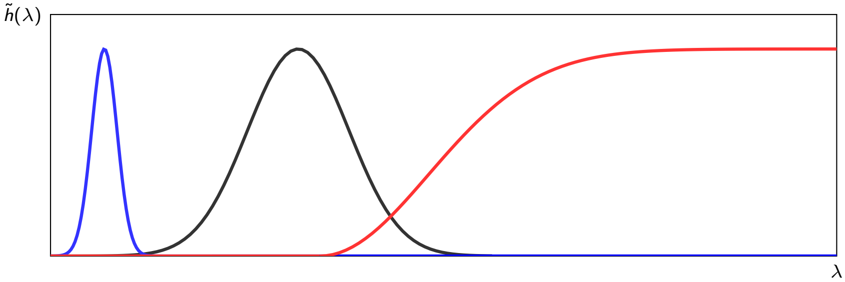
- In last bound, add and subtract $\mathbf{H}_\ell(\hat{\mathbf{S}})\mathbf{x}_{\ell-1}$. Triangle inequality. Submultiplicative property of norms

$$\left\|\mathbf{x}_\ell - \hat{\mathbf{x}}_\ell\right\| \leq \left\|\mathbf{H}_\ell(\mathbf{S})\mathbf{x}_{\ell-1} - \mathbf{H}_\ell(\hat{\mathbf{S}})\hat{\mathbf{x}}_{\ell-1} + \mathbf{H}_\ell(\hat{\mathbf{S}})\mathbf{x}_{\ell-1} - \mathbf{H}_\ell(\hat{\mathbf{S}})\mathbf{x}_{\ell-1}\right\|$$

$$\leq \left\|\mathbf{H}_\ell(\mathbf{S}) - \mathbf{H}_\ell(\hat{\mathbf{S}})\right\| \times \left\|\mathbf{x}_{\ell-1}\right\| + \left\|\mathbf{H}_\ell(\hat{\mathbf{S}})\right\| \times \left\|\mathbf{x}_{\ell-1} - \hat{\mathbf{x}}_{\ell-1}\right\|$$

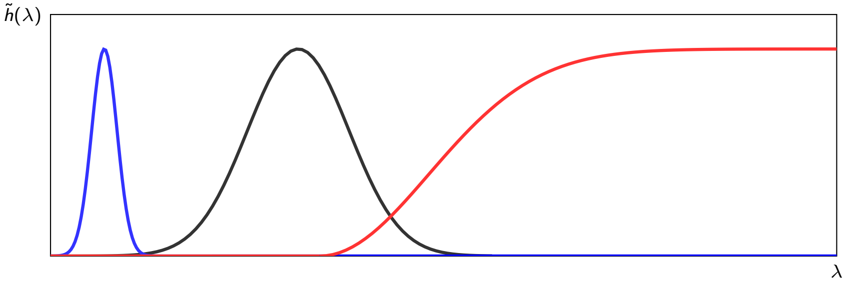
- Since filters are normalized $\Rightarrow$ Filter norm $\left\|\mathbf{H}_\ell(\hat{\mathbf{S}})\right\| = 1$. Signal norm $\Rightarrow \left\|\mathbf{x}_{\ell-1}\right\| \leq 1$
- The theorem on stability of filters to scaling holds $\Rightarrow \left\|\mathbf{H}_\ell(\mathbf{S}) - \mathbf{H}_\ell(\hat{\mathbf{S}})\right\| \leq \epsilon C + \mathcal{O}(\epsilon^2)$
- Put all bounds together $\Rightarrow \left\|\mathbf{x}_\ell - \hat{\mathbf{x}}_\ell\right\| \leq \epsilon C \times 1 + 1 \times \left\|\mathbf{x}_{\ell-1} - \hat{\mathbf{x}}_{\ell-1}\right\| + \mathcal{O}(\epsilon^2)$
- Apply recursively from Layer L back to Layer 1. The L factor appears ■

- GNNs have the same stability properties of graph filters. They need integral Lipschitz filters.
- Integral Lipschitz filters have to be flat at high frequencies. $\Rightarrow$ They can't discriminate
- It is impossible to separate signals with high frequency features and be stable to deformations



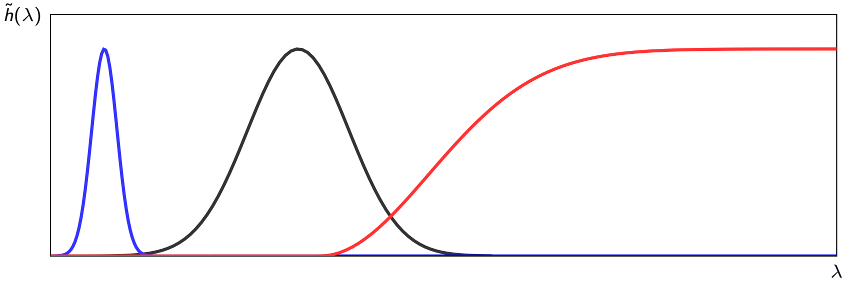
The Stability / Discriminability Tradeoff of GNNs

- GNNs have the same stability properties of graph filters. They need integral Lipschitz filters.
- On the flip side, integral Lipschitz filter can be very sharp at low frequencies
- We can be very discriminative at low frequencies. And at the same very stable to deformations



The Stability / Discriminability Tradeoff of GNNs

- GNNs use low-pass nonlinearities to demodulate high frequencies into low frequencies
- Where they can be discriminated sharply with a stable filter at the next layer
- Thus, they can be stable and discriminative. Something that linear graph filters can't be



Additive Perturbations of Graph Filters

- We define additive perturbations of the graph support

A Graph Filter can be Perturbed in Three Ways but Only One is Interesting

- Graph filter $\mathbf{H}(\mathbf{S})$ is a polynomial on shift operator $\mathbf{S}$ with coefficients h_k. Outputs given by

$$\mathbf{H}(\mathbf{S})\,\mathbf{x} = \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x}$$

- Perturbations of the input $\Rightarrow$ The filter is linear in $\mathbf{x}$. Scale error by filter's norm.
- Perturbations of the coefficients $\Rightarrow$ Filter is linear in h_k. Plus, h_k is a design parameter.
- Perturbations of the shift operator $\mathbf{S}$ $\Rightarrow$ It is not easy (nonlinear). And it is necessary.

$\Rightarrow$ The graph is estimated (recommendation systems). The graph changes (distributed systems)

$\Rightarrow$ Quasi-symmetries in graphs that are quasi-invariant to permutations

- Apply the same filter $\mathbf{h}$ to the same signal $\mathbf{x}$ on different graphs shift operators $\mathbf{S}$ and $\hat{\mathbf{S}}$

$$\mathbf{H}(\mathbf{S})\,\mathbf{x} \;=\; \sum_{k=0}^{K-1} h_k \mathbf{S}^k \mathbf{x} \qquad\qquad \mathbf{H}(\hat{\mathbf{S}})\,\mathbf{x} \;=\; \sum_{k=0}^{K-1} h_k \hat{\mathbf{S}}^k \mathbf{x}$$

- Filter $\mathbf{H}(\mathbf{S})\,\mathbf{x}$ $\Rightarrow$ Coefficients h_k. Input signal $\mathbf{x}$. Instantiated on shift $\mathbf{S}$
- Filter $\mathbf{H}(\hat{\mathbf{S}})\,\hat{\mathbf{x}}$ $\Rightarrow$ Same Coefficients h_k. Same Input signal $\mathbf{x}$. Instantiated on perturbed shift $\hat{\mathbf{S}}$

- We investigated scalings $\hat{\mathbf{S}} = (1+\epsilon)\mathbf{S}$ are an example. But we are after more generic models.

Additive Perturbation

- Additive perturbation model $\Rightarrow \hat{\mathbf{S}} = \mathbf{S} + \mathbf{E}$.

- Error matrix $\mathbf{E} = \hat{\mathbf{S}} - \mathbf{S}$ exists for any pair $\mathbf{S}$, $\hat{\mathbf{S}}$. $\Rightarrow$ It's norm $\|\mathbf{E}\|$ quantifies their difference

- A flaw $\Rightarrow$ Graphs $\mathbf{S}$ and $\hat{\mathbf{S}} = \mathbf{P}^T\mathbf{S}\mathbf{P}$ are the same (relabeling). Yet we may not have $\|\mathbf{E}\| = 0$.

- We know better $\Rightarrow$ Operator distances modulo permutation $\big\| \hat{\mathbf{S}} - \mathbf{S} \big\|_{\mathcal{P}} = \min_{\mathcal{P}} \big\| \hat{\mathbf{S}}\mathbf{P}^T - \mathbf{P}^T\mathbf{S} \big\|$

- We need a concrete handle on the error matrix. Start from set of symmetric error matrices

$$\mathcal{E}(\mathbf{S}, \hat{\mathbf{S}}) = \left\{ \tilde{\mathbf{E}} : \mathbf{P}^T \hat{\mathbf{S}} \mathbf{P} = \mathbf{S} + \tilde{\mathbf{E}}, \quad \mathbf{P} \in \mathcal{P} \right\}$$

- For each permutation $\mathbf{P} \in \mathcal{P}$ we have a different error matrix $\tilde{\mathbf{E}} = \mathbf{P}^T \hat{\mathbf{S}} \mathbf{P} - \mathbf{S}$ in the set $\mathcal{E}(\mathbf{S}, \hat{\mathbf{S}})$

- Error matrix modulo permutation is the one with smallest norm $\Rightarrow \mathbf{E} = \underset{\tilde{\mathbf{E}} \in \mathcal{E}(\mathbf{S}, \hat{\mathbf{S}})}{\operatorname{argmin}} \|\tilde{\mathbf{E}}\|$

- Rewrite the distance modulo permutation as $\Rightarrow d(\mathbf{S}, \hat{\mathbf{S}}) = \|\mathbf{E}\| = \min_{\tilde{\mathbf{E}} \in \mathcal{E}(\mathbf{S}, \hat{\mathbf{S}})} \|\tilde{\mathbf{E}}\|$

- Error norm $\|\mathbf{E}\| = d(\mathbf{S}, \hat{\mathbf{S}})$ measures how far $\mathbf{S}$ and $\hat{\mathbf{S}}$ are from being permutations of each other

Penn
UNIVERSITY of PENNSYLVANIA

- Consider eigenvector decompositions of the shift $\mathbf{S} = \mathbf{V}\Lambda\mathbf{V}^H$ and the error $\mathbf{E} = \mathbf{U}\mathbf{M}\mathbf{U}^H$

- Define the eigenvector misalignment between the shift operator $\mathbf{S}$ and the error matrix $\mathbf{E}$ as

$$\delta \;=\; \Big(\|\mathbf{U}-\mathbf{V}\| + 1\Big)^2 - 1$$

- Since $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices $\|\mathbf{U}\| = \|\mathbf{V}\| = 1 \;\Rightarrow \delta \leq 8 \;=\; [(2+1)^2 - 1]$

$\Rightarrow$ The eigenvector misalignment δ is never large. It can be small. Depending on the error model.

Stability of Lipschitz Filters to Additive Perturbations

- We show that Lipschitz filters are stable to additive perturbations of the graph support.

Theorem (Lipschitz Filters are Stable to Additive Perturbations)

Consider graph filter $\mathbf{h}$ along with shift operators $\mathbf{S}$ and $\hat{\mathbf{S}}$ having n nodes. If it holds that:

(H1) Shift operators $\mathbf{S}$ and $\hat{\mathbf{S}}$ are related by $\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{E}$ with $\mathbf{P}$ a permutation matrix

(H2) The error matrix $\mathbf{E}$ has norm $\|\mathbf{E}\| = \epsilon$ and eigenvector misalignement δ relative to $\mathbf{S}$

(H3) The filter $\mathbf{h}$ is Lipschitz with constant C

Then, the operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded by

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq C\left(1 + \delta\sqrt{n}\right)\epsilon + \mathcal{O}(\epsilon^2).$$

Theorem (Lipschitz Filters are Stable to Additive Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ satisfies

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq C \left(1 + \delta\sqrt{n}\right) \epsilon + \mathcal{O}(\epsilon^2).$$

- If shifts $\mathbf{S}$ and $\hat{\mathbf{S}}$ are ϵ-close the filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ are ϵ-close. Modulo permutation
- Proportional to the Lipschitz constant of the filter's frequency response. Not integral Lipschitz
- Proportional to $(1 + \delta\sqrt{n})$. Not great for large graphs. Unless misalignement decreases with n.
- Growth with n is at most $(1 + 8\sqrt{n}) \geq (1 + \delta\sqrt{n})$. Because $\delta \leq 8$. Not that bad

Stability is Stronger than Continuity

Theorem (Lipschitz Filters are Stable to Additive Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ satisfies

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq C \left(1 + \delta\sqrt{n} \right) \epsilon + \mathcal{O}(\epsilon^2).$$

- Filter perturbations are first order Lipschitz continuous with respect to the perturbation's size ϵ

  $\Rightarrow$ With Lipschitz constant $\Rightarrow C \left(1 + \delta\sqrt{n} \right)$

- Stronger than plain continuity. Which would say "output changes are small if input changes are"

Theorem (Lipschitz Filters are Stable to Additive Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded by

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq C \left(1 + \delta\sqrt{n}\right) \epsilon + \mathcal{O}(\epsilon^2).$$

- Bound is universal for all graphs with a given number of nodes n. Bound depends on:
  - $\Rightarrow$ A property of the filter's frequency response. The filter's Lipschitz constant C
  - $\Rightarrow$ And properties of the perturbation $\mathbf{E}$. The eigenvector misalignement δ and the norm $\|\mathbf{E}\| = \epsilon$
- There is no constant in the bound that depends on the graph shift operator $\mathbf{S}$. Save for n.

Theorem (Lipschitz Filters are Stable to Additive Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ satisfies

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq C\left(1 + \delta\sqrt{n}\right)\epsilon + \mathcal{O}(\epsilon^2).$$

- The filter's Lipschitz constant C is a parameter that we can affect with judicious filter choice
- Discriminability / stability tradeoff. Larger C improves discriminability at the cost of stability

Theorem (Lipschitz Filters are Stable to Additive Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded by

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq C \left(1 + \delta\sqrt{n}\right) \epsilon + \mathcal{O}(\epsilon^2).$$

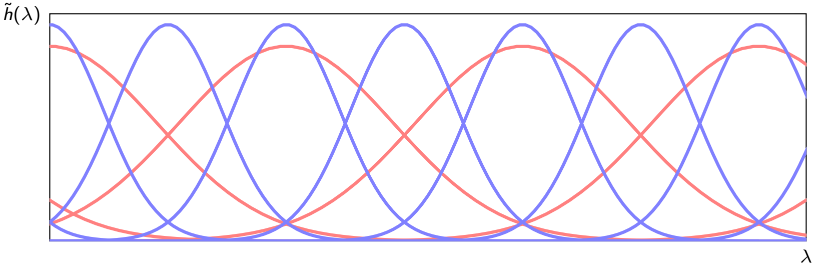
- Eigenvector misalignment δ is a property of the perturbation matrix. Independent of filter choice

  $\Rightarrow$ Not very relevant in studying stability / discriminability tradeoffs of different filters.

- Meaningless asymptotically on n. Don't know much about perturbations in the limit of large n

Lipschitz Filters are Good News

- Stability to additive perturbations requires Lipschitz filters. Not integral Lipschitz as with scalings
- Genuine stability / discriminability tradeoff $\Rightarrow$ Larger C tradeoffs stability for discriminability
- We can always discriminate, regardless of frequency, if we tolerate enough discriminability.



Relative Perturbations of Graph Filters

- Proved enticing stability properties with respect to additive perturbations. Alas, not meaningful

- We switch focus to relative perturbations. Which tie perturbations to the graph structure

- Additive perturbations are not meaningful

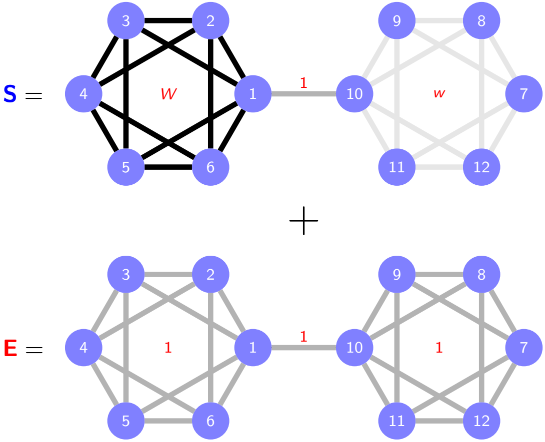
$$\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{E}$$

- With $w \ll 1 \ll W$.

  $\Rightarrow$ Is this perturbation small or large?

- Edges with small weights w can change a lot

  because other edges have large weights W



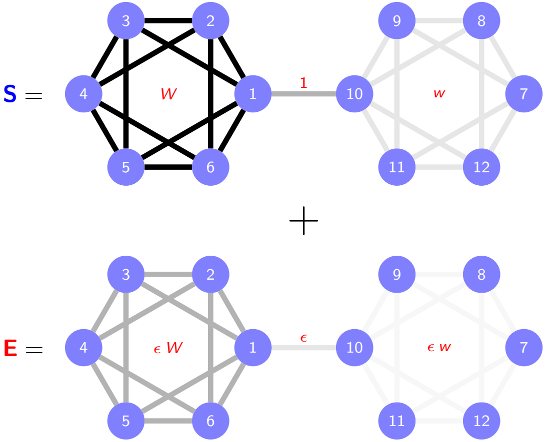
- Relative perturbations are more meaningful

$$\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{E} = \mathbf{S} + \epsilon\mathbf{I}\mathbf{S}$$

- With $w \ll 1 \ll W$ and $\epsilon \ll 1$

  $\Rightarrow$ Is this perturbation small or large?

- It's small. Edges with small weights change little. Edges with large weights change more



- Relative perturbation model $\Rightarrow \hat{\mathbf{S}} = \mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}$. We must account for permutations (relabeling)
- Set of relative error matrices modulo permutation. Matrices $\tilde{\mathbf{E}}$ are symmetric, $\tilde{\mathbf{E}} = \tilde{\mathbf{E}}^T$

$$\mathcal{E}(\mathbf{S}, \hat{\mathbf{S}}) = \left\{ \tilde{\mathbf{E}} \; : \; \mathbf{P}^T \hat{\mathbf{S}} \mathbf{P} \; = \; \mathbf{S} \; + \; \tilde{\mathbf{E}}\mathbf{S} \; + \; \mathbf{S}\tilde{\mathbf{E}} \; , \;\; \mathbf{P} \in \mathcal{P} \right\}$$

- Relative error matrix modulo permutation is the one with smallest norm $\Rightarrow \mathbf{E} \; = \; \underset{\tilde{\mathbf{E}} \in \mathcal{E}(\mathbf{S}, \hat{\mathbf{S}})}{\operatorname{argmin}} \|\tilde{\mathbf{E}}\|$
- Define relative distance modulo permutation as $\Rightarrow d(\mathbf{S}, \hat{\mathbf{S}}) \; = \; \|\mathbf{E}\| \; = \; \min_{\tilde{\mathbf{E}} \in \mathcal{E}(\mathbf{S}, \hat{\mathbf{S}})} \|\tilde{\mathbf{E}}\|$
- Norm $\|\mathbf{E}\| = d(\mathbf{S}, \hat{\mathbf{S}})$ is a relative measure of how far $\hat{\mathbf{S}}$ is from being a permutation of $\mathbf{S}$

- Relative perturbations tie changes in the edge weights to the local structure of the graph
- Compare edge weights in the given matrix $\mathbf{S}$ and the permuted version of the perturbations $\hat{\mathbf{S}}$

$$\begin{aligned}\left(\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P}\right)_{ij} &= S_{ij} + \left(\mathbf{ES}\right)_{ij} + \left(\mathbf{SE}\right)_{ij} \\ &= S_{ij} + \sum_{k\in n(j)} E_{ik}S_{kj} + \sum_{k\in n(i)} S_{ik}E_{kj}\end{aligned}$$

- Edge changes are proportional to the degree of the incident nodes. Scaled by entries of error matrix
- Parts of the graph with weaker connectivity see smaller changes than parts with stronger links
- In generic additive perturbations weights can change the same regardless of local connectivity

Stability of Integral Lipschitz Filters to Relative Perturbations

- We show that integral Lipschitz filters are stable to relative perturbations of the graph support.

Theorem (Integral Lipschitz Filters are Stable to Relative Perturbations)

Consider graph filter $\mathbf{h}$ along with shift operators $\mathbf{S}$ and $\hat{\mathbf{S}}$ having n nodes. If it holds that:

(H1) $\mathbf{S}$ and $\hat{\mathbf{S}}$ are related by $\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{ES} + \mathbf{SE}$ with $\mathbf{P}$ a permutation matrix

(H2) Error matrix has norm $\|\mathbf{E}\| = \epsilon$ and eigenvector misalignment constant δ relative to $\mathbf{S}$

(H3) The filter is integral Lipschitz with constant C

Then, the operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded by

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq 2C\left(1 + \delta\sqrt{n}\right)\epsilon + \mathcal{O}(\epsilon^2).$$

Theorem (Integral Lipschitz Filters are Stable to Relative Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded by

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq 2C \left(1 + \delta\sqrt{n} \right) \epsilon + \mathcal{O}(\epsilon^2).$$

- Save for the 2 factor, it is the same bound we have for the case of additive perturbations.
- The difference is in hypotheses (H1) and (H3). Hypothesis (H2) does not change

(H1) The perturbation is relative. $\Rightarrow \mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}$. Not additive.

(H3) The filter is integral Lipschitz with constant C. Not regular Lipschitz.

Theorem (Integral Lipschitz Filters are Stable to Relative Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded by

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq 2C \left(1 + \delta\sqrt{n}\right) \epsilon + \mathcal{O}(\epsilon^2).$$

- If $\mathbf{S}$ and $\hat{\mathbf{S}}$ are ϵ-close in relative terms, the filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ are ϵ-close. Modulo permutation
- Proportional to the integral Lipschitz constant of the filter's frequency response.
- Proportional to $(1 + \delta\sqrt{n})$. Not great for large graphs. Unless the misalignment decreases with n.

Stability is Stronger than Continuity

Theorem (Integral Lipschitz Filters are Stable to Relative Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ satisfies

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq 2C \left(1 + \delta\sqrt{n}\right) \epsilon + \mathcal{O}(\epsilon^2).$$

- Filter perturbations are first order Lipschitz continuous with respect to the perturbation's size ϵ

  $\Rightarrow$ With Lipschitz constant $\Rightarrow 2C \left(1 + \delta\sqrt{n}\right)$

- Stronger than plain continuity. Which would say "output changes are small if input changes are"

- Input perturbation measure is relative $\Rightarrow$ Norm $\|\mathbf{E}\| = \epsilon$ in mulitplicative perturbation $\mathbf{ES} + \mathbf{SE}$

Theorem (Integral Lipschitz Filters are Stable to Relative Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded by

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq 2C \left(1 + \delta\sqrt{n}\right) \epsilon + \mathcal{O}(\epsilon^2).$$

- Bound is universal for all graphs with a given number of nodes n. Bound depends on:
  - $\Rightarrow$ A property of the filter's frequency response. The filter's integral Lipschitz constant C
  - $\Rightarrow$ And properties of the perturbation $\mathbf{E}$. The eigenvector misalignement δ and the norm $\|\mathbf{E}\| = \epsilon$
- There is no constant in the bound that depends on the graph shift operator $\mathbf{S}$. Save for n.

Theorem (Integral Lipschitz Filters are Stable to Relative Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded by

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq 2C \left(1 + \delta\sqrt{n}\right) \epsilon + \mathcal{O}(\epsilon^2).$$

- Eigenvector misalignment δ is a property of the perturbation matrix. Independent of filter choice
- Meaningless asymptotically on n. Growth is not terrible. It is at most $1 + 8\sqrt{n}$

Filter's are Required to be Integral Lipschitz

Theorem (Integral Lipschitz Filters are Stable to Relative Perturbations)

The operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded by

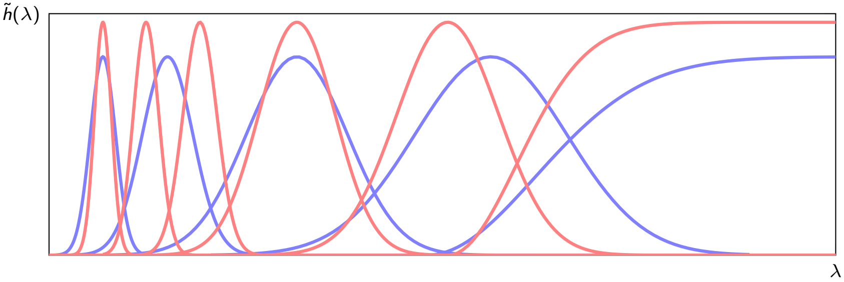
$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\|_{\mathcal{P}} \leq 2C \left(1 + \delta\sqrt{n}\right) \epsilon + \mathcal{O}(\epsilon^2).$$

- Bound depends on integral Lipschitz constant C. Very different from Lipschitz constant
- Can decrease C to increase stability. But effect on Discriminability depends on the frequency.

$\Rightarrow$ Discriminative at low frequencies regardless of C

$\Rightarrow$ Non-discriminative at high frequencies regardless of C

- Stability to relative perturbations requires integral Lipschitz filters. As in the case of dilations
- No stability vs discriminability tradeoff $\Rightarrow$ Stability and discriminability are incompatible
- No discriminability for large λ. Regardless of how much instability we tolerate by increasing C.



Stability Properties of Graph Neural Networks

- The stability properties we studied for graph filters are inherited by GNNs

- We proved that integral Lipschitz filters are stable to dilations of the shift operator

Theorem (Integral Lipschitz Graph Filters are Stable to Scaling)

Given graph shift operators $\mathbf{S}$ and $\hat{\mathbf{S}} = (1+\epsilon)\,\mathbf{S}$ and an integral Lipschitz filter with constant C.

The operator norm difference between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ is bounded as

$$\left\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \right\| \leq C\epsilon + \mathcal{O}(\epsilon^2).$$

- And that GNNs with integral Lipschitz layers inherit the stability of the filters to these dilations

Theorem (Integral Lipschitz GNNs are Stable to Scaling)

Given shift operators $\mathbf{S}$ and $\hat{\mathbf{S}} = (1+\epsilon)\,\mathbf{S}$ and a GNN operator $\Phi(\cdot;\mathbf{S},\mathcal{H})$ with L single-feature layers. The filters at each layer have unit operator norms and are integral Lipschitz with constant C. The nonlinearity σ is normalized Lipschitz. Then

$$\left\| \Phi(\cdot;\hat{\mathbf{S}},\mathcal{H}) - \Phi(\cdot;\mathbf{S},\mathcal{H}) \right\| \leq C\,L\,\epsilon + \mathcal{O}(\epsilon^2).$$

- The proof has nothing that is specific to dilations

  $\Rightarrow$ Any stability property that a class of graph filters has is inherited to a respective GNN

Theorem (Integral Lipschitz GNNs are Stable to Scaling)

Given shift operators $\mathbf{S}$ and $\hat{\mathbf{S}} = (1+\epsilon)\,\mathbf{S}$ and a GNN operator $\Phi(\cdot\,; \mathbf{S}, \mathcal{H})$ with L single-feature layers. The filters at each layer have unit operator norms and are integral Lipschitz with constant C. The nonlinearity σ is normalized Lipschitz. Then

$$\left\| \Phi(\cdot\,; \hat{\mathbf{S}}, \mathcal{H}) - \Phi(\cdot\,; \mathbf{S}, \mathcal{H}) \right\| \leq C\, L\, \epsilon + \mathcal{O}(\epsilon^2).$$

GNNs Inherit Any Stability Properties that Filters May Have

- Lipschitz filters are stable to additive deformations of the shift operator

  $\Rightarrow$ GNNs with Lipschitz layers are stable to additive deformations of the shift operator

- Integral Lipschitz filters are stable to relative deformations of the shift operator

  $\Rightarrow$ GNNs with integral Lipschitz layers are stable to relative deformations of the shift operator

Normalizations

- Reminders and precision are redundant but not unnecessary. Normalize filters and nonlinearities.

- At each layer of the GNN, the filters have unit operator norm $\Rightarrow \| \mathbf{H}_\ell(\mathbf{S}) \| = 1$

  $\Rightarrow$ Easy to achieve with scaling $\Rightarrow$ Equivalent to $\max_\lambda \tilde{h}_\ell(\lambda) = 1$

- The nonlinearity σ is Lipschitz and normalized so that $\Rightarrow \| \sigma(\mathbf{x}_2) - \sigma(\mathbf{x}_1) \| \leq \| \mathbf{x}_2 - \mathbf{x}_1 \|$

  $\Rightarrow$ Easy to achieve with scaling. True of ReLU, hyperbolic tangent, and absolute value

- Joining both assumptions $\Rightarrow$ If input energy is $\|\mathbf{x}\| \leq 1$, all layer outputs have energy $\|\mathbf{x}_\ell\| \leq 1$

Theorem (GNN Stability to Additive Perturbations)

Consider a GNN operator $\mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H})$ along with shifts operators $\mathbf{S}$ and $\hat{\mathbf{S}}$ having n nodes. If:

- **(H1)** Shift operators are related by $\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{E}$. With $\mathbf{P}$ a permutation matrix
- **(H2)** The error matrix $\mathbf{E}$ has norm $\|\mathbf{E}\| = \epsilon$ and eigenvector misalignement δ relative to $\mathbf{S}$
- **(H3)** The GNN has L single-feature layers with Lipschitz filters with constant C
- **(H4)** Filters have unit operator norm and the nonlinearity is normalized Lipschitz

Then, the operator distance modulo permutation between $\mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H})$ and $\mathbf{\Phi}(\cdot; \hat{\mathbf{S}}, \mathcal{H})$ is bounded by

$$\left\| \mathbf{\Phi}(\cdot; \hat{\mathbf{S}}, \mathcal{H}) - \mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H}) \right\|_{\mathcal{P}} \leq C\left(1 + \delta\sqrt{n}\right) L\epsilon + \mathcal{O}(\epsilon^2).$$

Theorem (GNN Stability to Additive Perturbations)

The operator distance modulo permutation between $\mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H})$ and $\mathbf{\Phi}(\cdot; \hat{\mathbf{S}}, \mathcal{H})$ is bounded by

$$\left\| \mathbf{\Phi}(\cdot; \hat{\mathbf{S}}, \mathbf{h}) - \mathbf{\Phi}(\cdot; \mathbf{S}, \mathbf{h}) \right\|_{\mathcal{P}} \leq C \left(1 + \delta\sqrt{n}\right) L \epsilon + \mathcal{O}(\epsilon^2).$$

- It is essentially the same bound we have for the case of Lipschitz filters. Propagated over L layers
- A GNN in which layers are made up of Lipschitz inherits the stability of the Lipschitz filter class
- The nonlinearity is pointwise $\Rightarrow$ Graph deformations have no effect on its action

Theorem (GNN Stability to Relative Perturbations)

Consider a GNN operator $\mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H})$ along with shifts operators $\mathbf{S}$ and $\hat{\mathbf{S}}$ having n nodes. If:

- **(H1)** Shift operators are related by $\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{ES} + \mathbf{SE}$ with $\mathbf{P}$ a permutation matrix
- **(H2)** The error matrix $\mathbf{E}$ has norm $\|\mathbf{E}\| = \epsilon$ and eigenvector misalignement δ relative to $\mathbf{S}$
- **(H3)** The GNN has L single-feature layers with integral Lipschitz filters with constant C
- **(H4)** Filters have unit operator norm and the nonlinearity is normalized Lipschitz

Then, the operator distance modulo permutation between $\mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H})$ and $\mathbf{\Phi}(\cdot; \hat{\mathbf{S}}, \mathcal{H})$ is bounded by

$$\left\| \mathbf{\Phi}(\cdot; \hat{\mathbf{S}}, \mathcal{H}) - \mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H}) \right\|_{\mathcal{P}} \leq 2C \left(1 + \delta\sqrt{n}\right) L\epsilon + \mathcal{O}(\epsilon^2).$$

Theorem (Single Feature GNN Stability to Relative Perturbations)

The operator distance modulo permutation between $\mathbf{\Phi}(\cdot;\mathbf{S},\mathcal{H})$ and $\mathbf{\Phi}(\cdot;\hat{\mathbf{S}},\mathcal{H})$ is bounded by

$$\left\| \mathbf{\Phi}(\cdot;\hat{\mathbf{S}},\mathbf{h}) - \mathbf{\Phi}(\cdot;\mathbf{S},\mathbf{h}) \right\|_{\mathcal{P}} \leq 2C \left(1 + \delta\sqrt{n}\right) L\epsilon + \mathcal{O}(\epsilon^2).$$

- It is essentially the same bound we have for integral Lipschitz filters. Propagated over L layers
- A GNN in which layers are integral Lipschitz inherits the stability of integral Lipschitz filters
- The nonlinearity is pointwise $\Rightarrow$ Graph deformations have no effect on its action

GNNs Inherit the Stability Properties of Graph Filters

- Let's do the proof for relative perturbations and integral Lipschitz filters.
- But this time we pay attention to the fact that steps apply to any stability claim on any filter class.
- And take the chance to discuss how GNNs inherit their stability properties from graph filters

Theorem (GNN Stability to Relative Perturbations)

Consider a GNN operator $\mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H})$ along with shifts operators $\mathbf{S}$ and $\hat{\mathbf{S}}$ having n nodes. If:

- **(H1)** Shift operators are related by $\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}$ with $\mathbf{P}$ a permutation matrix
- **(H2)** The error matrix $\mathbf{E}$ has norm $\|\mathbf{E}\| = \epsilon$ and eigenvector misalignement δ relative to $\mathbf{S}$
- **(H3)** The GNN has L single-feature layers with integral Lipschitz filters with constant C
- **(H4)** Filters have unit operator norm and the nonlinearity is normalized Lipschitz

Then, the operator distance modulo permutation between $\mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H})$ and $\mathbf{\Phi}(\cdot; \hat{\mathbf{S}}, \mathcal{H})$ is bounded by

$$\left\| \mathbf{\Phi}(\cdot; \hat{\mathbf{S}}, \mathcal{H}) - \mathbf{\Phi}(\cdot; \mathbf{S}, \mathcal{H}) \right\|_{\mathcal{P}} \leq 2C\left(1 + \delta\sqrt{n}\right) L\epsilon + \mathcal{O}(\epsilon^2).$$

Proof: Let $\mathbf{x}_\ell$ be the Layer ℓ output of GNN $\Phi(\mathbf{x};\mathbf{S},\mathcal{H})$. Input signal $\mathbf{x}$ with $\|\mathbf{x}\|=1$

Let $\hat{\mathbf{x}}_\ell$ be the Layer ℓ output of GNN $\Phi(\mathbf{x};\hat{\mathbf{S}},\mathcal{H})$. Input signal $\mathbf{x}$ with $\|\mathbf{x}\|=1$

- Layer ℓ is a perceptron with filter $\mathbf{H}_\ell$ $\Rightarrow$ $\Big\|\hat{\mathbf{x}}_\ell-\mathbf{x}_\ell\Big\| = \Bigg\|\sigma\Big[\mathbf{H}_\ell(\hat{\mathbf{S}})\hat{\mathbf{x}}_{\ell-1}\Big]-\sigma\Big[\mathbf{H}_\ell(\mathbf{S})\mathbf{x}_{\ell-1}\Big]\Bigg\|$
- Nonlinearity is normalized Lipschitz $\Rightarrow$ $\Big\|\hat{\mathbf{x}}_\ell-\mathbf{x}_\ell\Big\| \leq \Big\|\mathbf{H}_\ell(\hat{\mathbf{S}})\hat{\mathbf{x}}_{\ell-1}-\mathbf{H}_\ell(\mathbf{S})\mathbf{x}_{\ell-1}\Big\|$
- This is the critical step of the proof. The rest of the proof is just algebra.

- In last bound, add and subtract $\mathbf{H}_\ell(\hat{\mathbf{S}})\mathbf{x}_{\ell-1}$. Triangle inequality. Submultiplicative property of norms

$$\left\| \hat{\mathbf{x}}_\ell - \mathbf{x}_\ell \right\| \leq \left\| \mathbf{H}_\ell(\hat{\mathbf{S}})\hat{\mathbf{x}}_{\ell-1} - \mathbf{H}_\ell(\mathbf{S})\mathbf{x}_{\ell-1} + \mathbf{H}_\ell(\hat{\mathbf{S}})\mathbf{x}_{\ell-1} - \mathbf{H}_\ell(\hat{\mathbf{S}})\mathbf{x}_{\ell-1} \right\|$$

$$\leq \left\| \mathbf{H}_\ell(\hat{\mathbf{S}}) - \mathbf{H}_\ell(\mathbf{S}) \right\| \times \left\| \mathbf{x}_{\ell-1} \right\| + \left\| \mathbf{H}_\ell(\hat{\mathbf{S}}) \right\| \times \left\| \hat{\mathbf{x}}_{\ell-1} - \mathbf{x}_{\ell-1} \right\|$$

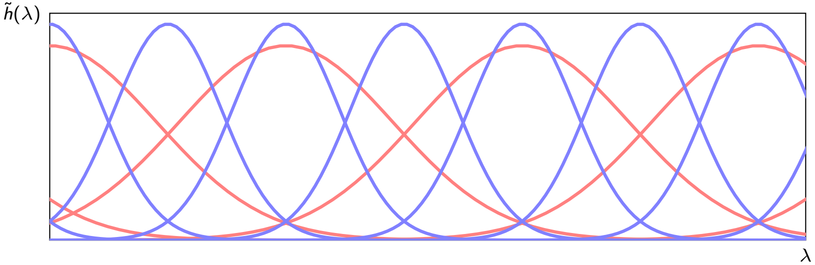
- Since filters are normalized $\Rightarrow$ Filter norm $\left\| \mathbf{H}_\ell(\hat{\mathbf{S}}) \right\| = 1$. Signal norm $\Rightarrow \left\| \mathbf{x}_{\ell-1} \right\| \leq 1$
- Relative perturbations and integral Lipschitz $\Rightarrow \left\| \mathbf{H}_\ell(\hat{\mathbf{S}}) - \mathbf{H}_\ell(\mathbf{S}) \right\| \leq 2C\left(1+\delta\sqrt{n}\right)\epsilon + \mathcal{O}(\epsilon^2)$
- Put all bounds together $\Rightarrow \left\| \hat{\mathbf{x}}_\ell - \mathbf{x}_\ell \right\| \leq 2C\left(1+\delta\sqrt{n}\right)\epsilon \times 1 + 1 \times \left\| \hat{\mathbf{x}}_{\ell-1} - \mathbf{x}_{\ell-1} \right\| + \mathcal{O}(\epsilon^2)$
- Apply recursively from Layer L back to Layer 1. The L factor appears ∎

GNNs Inherit the Stability of Graph Filters

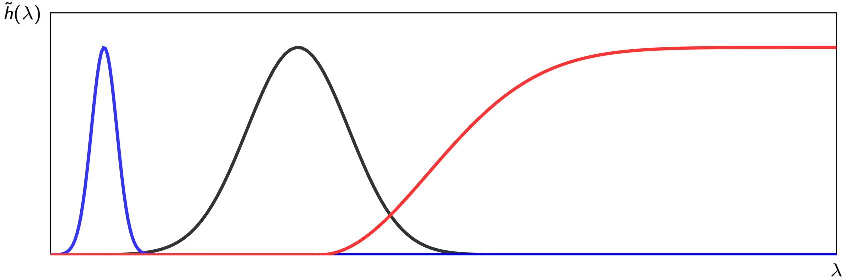
Since Stability is inherited from graph filters, mutatis mutandis, the same observations hold here.

- We claim stability. Which is stronger than continuity.
- The stability bounds are universal for all graphs with a given number of nodes
- Bounds depend on filter's Lipschitz constant C and the number of layers L. Which we control.
- And the eigenvector misalignment constant. Which we don't control. Depends on the perturbation.

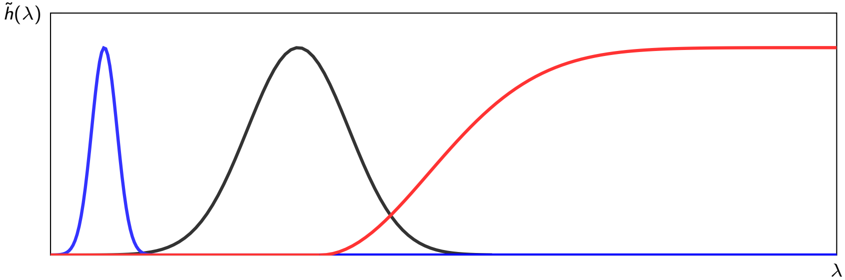
- GNNs whose layers are made up of Lipschitz graph filters are stable to additive deformations
- This is good news $\Rightarrow$ We have a genuine stability vs discriminability tradeoff
- Alas, a bit of a mirage $\Rightarrow$ Graph perturbations are more naturally measured in relative tems



- Meaningful stability claims with respect to relative perturbations require integral Lipschitz filters.
- Integral Lipschitz filters have to be flat at high frequencies. $\Rightarrow$ They can't discriminate
- It is impossible to separate signals with high frequency features and be stable to deformations



- Meaningful stability claims with respect to relative perturbations require integral Lipschitz filters.
- On the flip side, integral Lipschitz filter can be very sharp at low frequencies
- We can be very discriminative at low frequencies. And at the same very stable to deformations



- GNNs use low-pass nonlinearities to demodulate high frequencies into low frequencies
- Where they can be discriminated sharply with a stable filter at the next layer
- Thus, they can be stable and discriminative. Something that linear graph filters can't be