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 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 $\left\| \hat{\mathbf{S}} - \mathbf{S} \right\|_{\mathcal{P}} = \min_{\mathcal{P}} \left\| \hat{\mathbf{S}}\mathbf{P}^T - \mathbf{P}^T\mathbf{S} \right\|$

- 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

Eigenvector Misalignment Constant

- 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

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

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

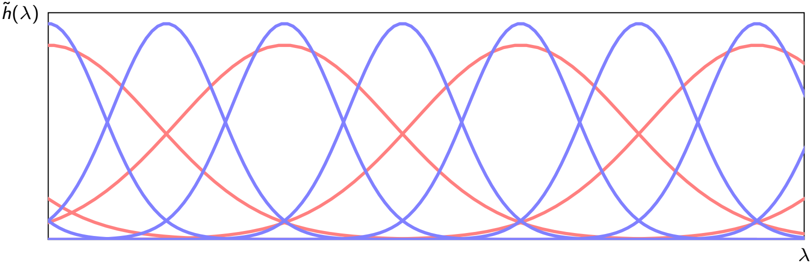
- Eigenvector misalignment δ is a property of the perturbation matrix. Independent of filter choice

  $\Rightarrow$ 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

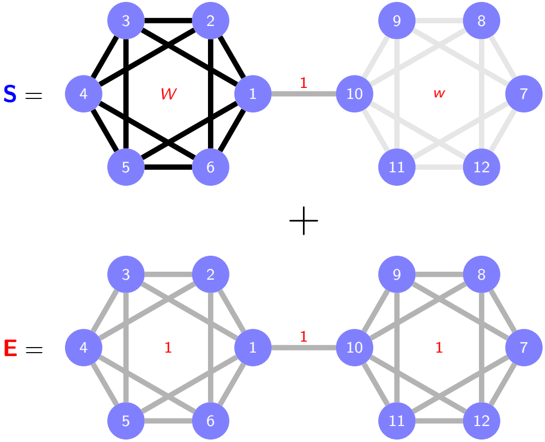
$$\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{E}$$

- 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

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

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

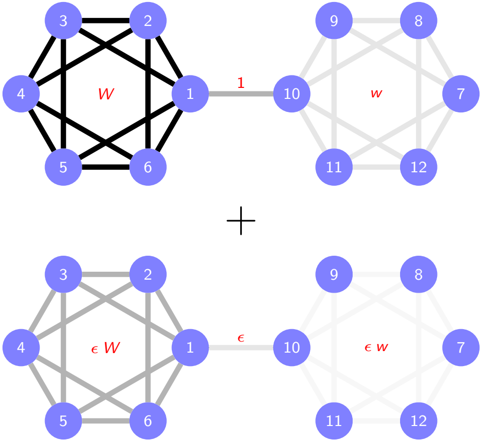
  $\Rightarrow$ Is this perturbation small or large?

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

$\mathbf{S} =$

$+$

$\mathbf{E} =$



- 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 $\hat{\mathbf{S}}$

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

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

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 $1 + 8\sqrt{n}$

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

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

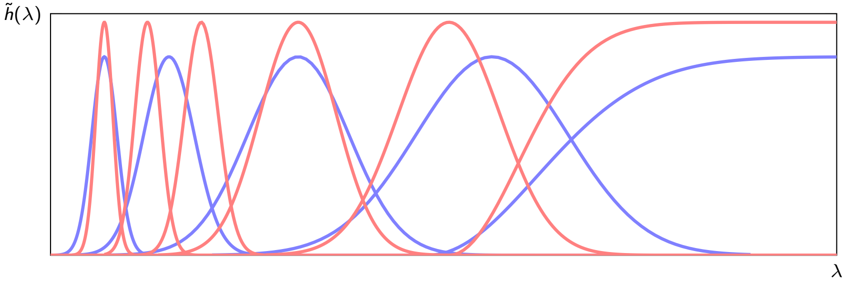
- Bound depends on integral Lipschitz constant C. Very different from Lipschitz constant
- Can decrease C to increase stability. But effect on Discriminability depends on the frequency.

$\Rightarrow$ Discriminative at low frequencies regardless of C

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

Integral Lipschitz Filters are Not Good News

- 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

- 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 $\mathbf{P}^T\hat{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{ES} + \mathbf{SE}$ 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

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

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

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

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

- Since filters are normalized $\Rightarrow$ Filter norm $\Big\| \mathbf{H}_\ell(\hat{\mathbf{S}}) \Big\| = 1$. Signal norm $\Rightarrow \Big\| \mathbf{x}_{\ell-1} \Big\| \leq 1$

- Relative perturbations and integral Lipschitz $\Rightarrow \Big\| \mathbf{H}_\ell(\hat{\mathbf{S}}) - \mathbf{H}_\ell(\mathbf{S}) \Big\| \leq 2C\big(1 + \delta\sqrt{n}\big)\epsilon + \mathcal{O}(\epsilon^2)$

- Put all bounds together $\Rightarrow \Big\| \hat{\mathbf{x}}_\ell - \mathbf{x}_\ell \Big\| \leq 2C\big(1 + \delta\sqrt{n}\big)\epsilon \times 1 + 1 \times \Big\| \hat{\mathbf{x}}_{\ell-1} - \mathbf{x}_{\ell-1} \Big\| + \mathcal{O}(\epsilon^2)$

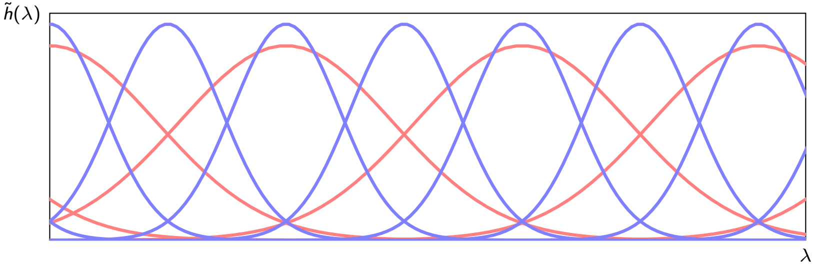
- Apply recursively from Layer L back to Layer 1. The L factor appears ■

GNNs Inherit the Stability of Graph Filters

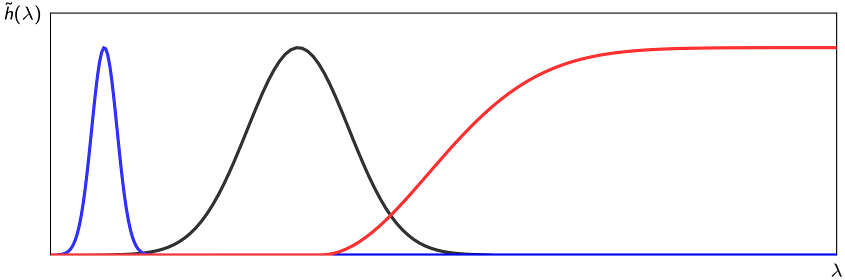
Since Stability is inherited from graph filters, mutatis mutandis, the same observations hold here.

- 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.

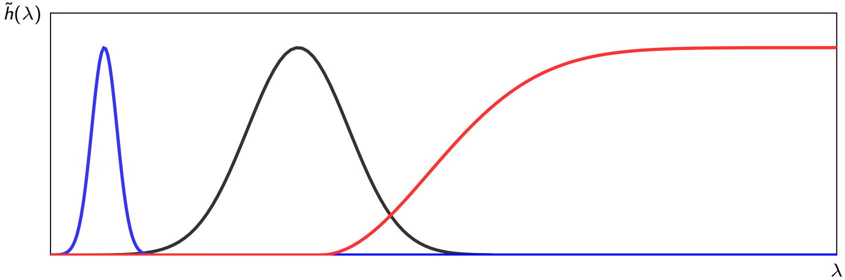
- GNNs whose layers are made up of Lipschitz graph filters are stable to additive deformations
- This is good news $\Rightarrow$ We have a genuine stability vs discriminability tradeoff
- Alas, a bit of a mirage $\Rightarrow$ Graph perturbations are more naturally measured in relative tems



- 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