# Stability of GRNNs

Alejandro Ribeiro, Luana Ruiz, and Fernando Gama

## 1 Preliminary results, definitions and assumptions

The equations that describe the GRNN are represented by the following equations

$$\mathbf{z}_t = \sigma\big(\mathbf{A}(\mathbf{S})\mathbf{x}_t + \mathbf{B}(\mathbf{S})\mathbf{z}_{t-1}\big) \tag{1}$$

$$\hat{\mathbf{y}}_t = \rho\left(\mathbf{C}(\mathbf{S})\mathbf{z}_t\right) \tag{2}$$

### 1.1 Preliminary results and definitions

**Proposition 1** *Let $\mathbf{S}$ be a GSO and $\tilde{\mathbf{S}} = \mathbf{P}^{\mathsf{T}}\mathbf{S}\mathbf{P}$ be a permutation of this GSO, for some permutation matrix $\mathbf{P} \in \mathcal{P}$. Let $\mathbf{x}_t$ be a graph signal and $\tilde{\mathbf{x}}_t = \mathbf{P}^{\mathsf{T}}\mathbf{x}_t$ the permuted version of the signal. Then, it holds that*

$$\tilde{\mathbf{z}}_t = \sigma(\mathbf{A}(\tilde{\mathbf{S}})\tilde{\mathbf{x}}_t + \mathbf{B}(\tilde{\mathbf{S}})\tilde{\mathbf{z}}_{t-1}) = \mathbf{P}^{\mathsf{T}}\mathbf{z}_t \tag{3}$$

$$\tilde{\mathbf{y}}_t = \rho(\mathbf{C}(\tilde{\mathbf{S}})\tilde{\mathbf{z}}_t) = \mathbf{P}^{\mathsf{T}}\mathbf{y}_t \quad \textit{for all } t. \tag{4}$$

*where $\mathcal{P} = \big\{\mathbf{P} \in \{0,1\}^{N\times N} : \mathbf{P}\mathbf{1} = \mathbf{1}, \mathbf{P}^{\mathsf{T}}\mathbf{1} = \mathbf{1}\big\}$.*

**Proof:** Refer to Appendix 4.1. ∎

**Definition 1 (Relative perturbation matrices)** *Given GSOs $\mathbf{S}$ and $\tilde{\mathbf{S}}$, we define the set of relative perturbation matrices modulo permutation as*

$$\mathcal{E}(\mathbf{S}, \tilde{\mathbf{S}}) = \Big\{\mathbf{E} \in \mathbb{R}^{N\times N} : \mathbf{P}^{\mathsf{T}}\tilde{\mathbf{S}}\mathbf{P} = \mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}^{\mathsf{T}}, \mathbf{P} \in \mathcal{P}\Big\}. \tag{5}$$

We define the distance between two graphs described by $\mathbf{S}$ and $\tilde{\mathbf{S}}$ respectively as

$$d(\mathbf{S}, \tilde{\mathbf{S}}) = \min_{\mathbf{E} \in \mathcal{E}(\mathbf{S}, \hat{\mathbf{S}})} \|\mathbf{E}\|. \tag{6}$$

Notice that if $\tilde{\mathbf{S}}$ is a permutation of $\mathbf{S}$, then $d(\mathbf{S}, \tilde{\mathbf{S}}) = 0$.

**Definition 2 (Integral Lipschitz filters)** *We say that the filter $\mathbf{A}(\mathbf{S}) = \sum_{k=0}^{K-1} a_k \mathbf{S}^k$ is integral Lipschitz if there exists $C$ such that its frequency response $a(\lambda) = \sum_{k=0}^{K-1} a_k \lambda^k$, satisfies*

$$|a(\lambda_2) - a(\lambda_1)| \leq C \frac{|\lambda_2 - \lambda_1|}{|\lambda_1 + \lambda_2|/2} \tag{7}$$

*for all $\lambda_1, \lambda_2 \in \mathbb{R}$.*

Integral Lipschitz filters also satisfy $|\lambda a'(\lambda)| \leq C$, where $a'(\lambda)$ is the derivative of $a(\lambda)$.

## 1.2 Assumptions

Under the following assumptions, we prove that GRNNs built from integral Lipschitz filters are stable to relative perturbations in Theorem 1.

**Assumption 1** *The filters $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ of the GRNN (1)-(2) are integral Lipschitz [cf. (7)] with constants $C_{\mathbf{A}}$, $C_{\mathbf{B}}$ and $C_{\mathbf{C}}$ and normalized filter height $\|\mathbf{A}\| = \|\mathbf{B}\| = \|\mathbf{C}\| = 1$, respectively.*

**Assumption 2** *The pointwise nonlinearities $\sigma$ and $\rho$ (1)-(2) are normalized Lipschitz, i.e. $|\sigma(b) - \sigma(a)| \leq |b - a|$ for all $a, b \in \mathbb{R}$, and satisfy $\sigma(0) = \rho(0) = 0$.*

**Assumption 3** *The initial hidden state is identically zero, i.e. $\mathbf{z}_0 = \mathbf{0}$.*

**Assumption 4** *The inputs $\mathbf{x}_t$ satisfy $\|\mathbf{x}_t\| \leq \|\mathbf{x}\| = 1$ for every $t$.*

**Assumption 5** *We focus on single-feature GRNNs.*

## 2 Stability of the GRNNs

**Theorem 1 (Stability of GRNNs)** *Consider two graphs with $N$ nodes represented by the GSOs $\mathbf{S} = \mathbf{V\Lambda V}^{\mathsf{H}}$ and $\tilde{\mathbf{S}}$. Let $\mathbf{E} = \mathbf{UMU}^{\mathsf{H}} \in \mathcal{E}(\mathbf{S}, \tilde{\mathbf{S}})$ be a relative perturbation matrix [cf. (5)] such that [cf. (6)]*

$$d(\mathbf{S}, \tilde{\mathbf{S}}) \leq \|\mathbf{E}\| \leq \varepsilon. \tag{8}$$

*Let $\mathbf{y}_t$ and $\tilde{\mathbf{y}}_t$ be the outputs of GRNNs (1)-(2) running on $\mathbf{S}$ and $\tilde{\mathbf{S}}$ respectively, and satisfying AS1 through AS4. Then, it holds that*

$$\min_{\mathbf{P}\in\mathcal{P}} \|\mathbf{y}_t - \mathbf{P}^{\mathsf{T}}\tilde{\mathbf{y}}_t\| \leq C(1+\sqrt{N}\delta)(t^2+3t)\varepsilon \, + \mathcal{O}(\varepsilon^2) \tag{9}$$

*where $C$ is the maximum filter constant,*

$$C = \max\{C_{\mathbf{A}}, C_{\mathbf{B}}, C_{\mathbf{C}}\}$$

*and $\delta = (\|\mathbf{U}-\mathbf{V}\|+1)^2 - 1$ measures the eigenvector misalignment between the GSO $\mathbf{S}$ and the error matrix $\mathbf{E}$.*

## 3 Proof of stability of the GRNNs

**Lemma 1** *Let $\mathbf{S} = \mathbf{V\Lambda V}^{\mathsf{H}}$ and $\tilde{\mathbf{S}}$ be graph shift operators. Let $\mathbf{E} = \mathbf{UMU}^{\mathsf{H}} \in \mathcal{E}(\mathbf{S}, \tilde{\mathbf{S}})$ be a relative perturbation matrix [cf. Definition 1] whose norm is such that*

$$d(\mathbf{S}, \tilde{\mathbf{S}}) \leq \|\mathbf{E}\| \leq \varepsilon.$$

*For an integral Lipschitz filter [cf. Definition 2] with integral Lipschitz constant $C$, the operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\tilde{\mathbf{S}})$ satisfies*

$$\|\mathbf{H}(\mathbf{S}) - \mathbf{H}(\tilde{\mathbf{S}})\|_{\mathcal{P}} \leq 2C\left(1+\delta\sqrt{N}\right)\varepsilon + \mathcal{O}(\varepsilon^2) \tag{10}$$

*with $\delta := (\|\mathbf{U}-\mathbf{V}\|_2+1)^2 - 1$ standing for the eigenvector misalignment between shift operator $\mathbf{S}$ and error matrix $\mathbf{E}$.*

**Proof:** See [1, Theorem 3]. ∎

**Proof:** [Proof of Theorem 1]

Without loss of generality, assume $\mathbf{P} = \mathbf{I}$ in (6) and write $\tilde{\mathbf{S}} = \mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}^\top$, $\tilde{\mathbf{A}} = \mathbf{A}(\tilde{\mathbf{S}})$, $\tilde{\mathbf{B}} = \mathbf{B}(\tilde{\mathbf{S}})$ and $\tilde{\mathbf{C}} = \mathbf{C}(\tilde{\mathbf{S}})$. From (2), we can write

$$\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\| = \|\rho(\mathbf{C}\mathbf{z}_t) - \rho(\tilde{\mathbf{C}}\tilde{\mathbf{z}}_t)\| \leq \|\mathbf{C}\mathbf{z}_t - \tilde{\mathbf{C}}\tilde{\mathbf{z}}_t\| \tag{11}$$

since $\rho(\cdot)$ is normalized Lipschitz. Adding and subtracting $\mathbf{C}\tilde{\mathbf{z}}$ on the right-hand side of (11), and using both the triangle and Cauchy-Schwarz inequalities, we get

$$\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\| \leq \|\mathbf{C}\|\|\mathbf{z}_t - \tilde{\mathbf{z}}_t\| + \|\mathbf{C} - \tilde{\mathbf{C}}\|\|\tilde{\mathbf{z}}_t\|. \tag{12}$$

The norm of $\mathbf{C}$ is assumed bounded, and Lemma 1 gives a bound to $\|\mathbf{C} - \tilde{\mathbf{C}}\|$. Using (1), we can write

$$\begin{aligned}
\|\mathbf{z}_t - \tilde{\mathbf{z}}_t\| &= \|\sigma(\mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{z}_{t-1}) - \sigma(\tilde{\mathbf{A}}\mathbf{x}_t + \tilde{\mathbf{B}}\tilde{\mathbf{z}}_{t-1})\| \quad (13)\\
&\leq \|\mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{z}_{t-1} - (\tilde{\mathbf{A}}\mathbf{x}_t + \tilde{\mathbf{B}}\tilde{\mathbf{z}}_{t-1})\| \quad (14)\\
&\leq \|\mathbf{A} - \tilde{\mathbf{A}}\|\|\mathbf{x}_t\| + \|\mathbf{B}\mathbf{z}_{t-1} - \tilde{\mathbf{B}}\tilde{\mathbf{z}}_{t-1}\| \quad (15)
\end{aligned}$$

where the first inequality follows from the fact that $\sigma(\cdot)$ is also normalized Lipschitz and the second from the triangle and Cauchy-Schwarz inequalities respectively. The norm difference $\|\mathbf{A} - \tilde{\mathbf{A}}\|$ is bounded by Lemma 1 and $\|\mathbf{x}_t\| \leq \|\mathbf{x}\|$ for all $t$, so we move onto deriving a bound for the second summand of (15). We rewrite it as

$$\begin{aligned}
\|\mathbf{B}\mathbf{z}_{t-1} &+ \mathbf{B}\tilde{\mathbf{z}}_{t-1} - \mathbf{B}\tilde{\mathbf{z}}_{t-1} - \tilde{\mathbf{B}}\tilde{\mathbf{z}}_{t-1}\| \\
&\leq \|\mathbf{B}\|\|\mathbf{z}_{t-1} - \tilde{\mathbf{z}}_{t-1}\| + \|\mathbf{B} - \tilde{\mathbf{B}}\|\|\tilde{\mathbf{z}}_{t-1}\|
\end{aligned} \tag{16}$$

which results in a recurrence relationship between $\|\mathbf{z}_t - \tilde{\mathbf{z}}_t\|$ and $\|\mathbf{z}_{t-1} - \tilde{\mathbf{z}}_{t-1}\|$. Expanding this recurrence, we obtain

$$\begin{aligned}
\|\mathbf{z}_t - \tilde{\mathbf{z}}_t\| \leq &\sum_{i=0}^{t-1} \|\mathbf{B}\|^i \|\mathbf{A} - \tilde{\mathbf{A}}\|\|\mathbf{x}\| \\
&+ \|\mathbf{B}\|^t \|\mathbf{z}_0 - \tilde{\mathbf{z}}_0\| + \|\mathbf{B} - \tilde{\mathbf{B}}\| \sum_{i=1}^{t} \|\tilde{\mathbf{z}}_{t-i}\| \\
\leq &\sum_{i=0}^{t-1} \|\mathbf{B}\|^i \|\mathbf{A} - \tilde{\mathbf{A}}\|\|\mathbf{x}\| + \|\mathbf{B} - \tilde{\mathbf{B}}\| \sum_{i=0}^{t-1} \|\tilde{\mathbf{z}}_i\|
\end{aligned}$$

where the second inequality follows from $\mathbf{z}_0 = \tilde{\mathbf{z}}_0$. Now it suffices to bound $\|\mathbf{z}_i\|$ for any given $i > 0$. Writing $\mathbf{z}_t$ as in (1) and observing that, because $\sigma(\cdot)$ is normalized Lipschitz and $\sigma(0) = 0$, $|\sigma(x)| < |x|$, we can use the triangle and Cauchy-Schwarz inequalities to write

$$\|\mathbf{z}_i\| \leq \|\mathbf{A}\|\|\mathbf{x}_i\| + \|\mathbf{B}\|\|\mathbf{z}_{i-1}\| \leq \sum_{j=0}^{i-1} \|\mathbf{B}\|^j \|\mathbf{A}\|\|\mathbf{x}\| + \|\mathbf{B}\|^i \|\mathbf{z}_0\| \tag{17}$$

for $i > 0$. Substituting this in (16), we get

$$\begin{aligned} \|\mathbf{z}_t - \tilde{\mathbf{z}}_t\| \leq \|\mathbf{A} - \tilde{\mathbf{A}}\|\|\mathbf{x}\| \sum_{i=0}^{t-1} \|\mathbf{B}\|^i \\ + \|\mathbf{B} - \tilde{\mathbf{B}}\| \left( \|\tilde{\mathbf{A}}\|\|\mathbf{x}\| \sum_{i=1}^{t-1}\sum_{j=0}^{i-1} \|\tilde{\mathbf{B}}\|^j + \|\mathbf{z}_0\| \sum_{i=0}^{t-1} \|\tilde{\mathbf{B}}\|^i \right). \end{aligned} \tag{18}$$

Finally, substituting equations (16) and (18) in (12) gives

$$\begin{aligned} \|\mathbf{y}_t - \tilde{\mathbf{y}}_t\| \leq \|\mathbf{C}\| \Bigg[ \|\mathbf{A} - \tilde{\mathbf{A}}\|\|\mathbf{x}\| \sum_{i=0}^{t-1} \|\mathbf{B}\|^i \\ + \|\mathbf{B} - \tilde{\mathbf{B}}\| \left( \|\tilde{\mathbf{A}}\|\|\mathbf{x}\| \sum_{i=1}^{t-1}\sum_{j=0}^{i-1} \|\tilde{\mathbf{B}}\|^j + \|\mathbf{z}_0\| \sum_{i=0}^{t-1} \|\tilde{\mathbf{B}}\|^i \right) \Bigg] \\ + \|\mathbf{C} - \tilde{\mathbf{C}}\| \left[ \sum_{i=0}^{t-1} \|\tilde{\mathbf{B}}\|^i \|\tilde{\mathbf{A}}\|\|\mathbf{x}\| + \|\tilde{\mathbf{B}}\|^t \|\mathbf{z}_0\| \right]. \end{aligned}$$

This expression can be simplified by applying Lemma 1 to the norm differences $\|\mathbf{A} - \tilde{\mathbf{A}}\|$, $\|\mathbf{B} - \tilde{\mathbf{B}}\|$ and $\|\mathbf{C} - \tilde{\mathbf{C}}\|$, and by recalling that $\|\mathbf{A}\| = \|\mathbf{B}\| = \|\mathbf{C}\| = 1$, $\|\mathbf{x}\| = 1$ and $\mathbf{z}_0 = \mathbf{0}$. Denoting $C = \max\{C_{\mathbf{A}}, C_{\mathbf{B}}, C_{\mathbf{C}}\}$ the maximum filter Lipschitz constant, we recover (9) with $\mathbf{P} = \mathbf{I}$,

$$\|\mathbf{y}_t - \tilde{\mathbf{y}}_t\| \leq C(1 + \sqrt{N}\delta)(t^2 + 3t)\varepsilon \; + \mathcal{O}(\varepsilon^2) \tag{19}$$

which completes the proof. ■

# 4 Proofs of Auxiliary Results

## 4.1 Proof of Proposition 1

**Proof:** [Proof of Proposition 1] Since the permutation matrix $\mathbf{P} \in \mathcal{P}$ is orthogonal, we have $\mathbf{P}^{\mathsf{T}}\mathbf{P} = \mathbf{P}\mathbf{P}^{\mathsf{T}}$, which implies

$$\tilde{\mathbf{S}}^k = (\mathbf{P}^{\mathsf{T}}\mathbf{S}\mathbf{P})^k = \mathbf{P}^{\mathsf{T}}\mathbf{S}^k\mathbf{P}. \tag{20}$$

taking into account that

$$\mathbf{A}(\tilde{\mathbf{S}}) = \mathbf{P}^{\mathsf{T}}\mathbf{A}(\mathbf{S})\mathbf{P} \tag{21}$$

and applying $\mathbf{A}(\tilde{\mathbf{S}})$ to $\tilde{\mathbf{x}} = \mathbf{P}^{\mathsf{T}}\mathbf{x}$ yields

$$\mathbf{A}(\tilde{\mathbf{S}})\tilde{\mathbf{x}} = \mathbf{P}^{\mathsf{T}}\mathbf{A}(\mathbf{S})\mathbf{P}\mathbf{P}^{\mathsf{T}}\mathbf{x} = \mathbf{P}^{\mathsf{T}}\mathbf{A}(\mathbf{S})\mathbf{x}. \tag{22}$$

Graph convolutions are thus permutation equivariant. Using (1), we can then write $\tilde{\mathbf{z}}_t$ as

$$\tilde{\mathbf{z}}_t = \sigma(\mathbf{A}(\tilde{\mathbf{S}})\tilde{\mathbf{x}}_t + \mathbf{B}(\tilde{\mathbf{S}})\tilde{\mathbf{z}}_{t-1}) \tag{23}$$

$$= \sigma(\mathbf{P}^{\mathsf{T}}\mathbf{A}(\mathbf{S})\mathbf{x}_t + \mathbf{P}^{\mathsf{T}}\mathbf{B}(\mathbf{S})\mathbf{z}_{t-1}) \tag{24}$$

$$= \mathbf{P}^{\mathsf{T}}\sigma(\mathbf{A}(\mathbf{S})\mathbf{x}_t + \mathbf{B}(\mathbf{S})\mathbf{z}_{t-1}) = \mathbf{P}^{\mathsf{T}}\mathbf{z}_t \tag{25}$$

where the second-to-last equality follows from the fact that $\sigma$ is pointwise and hence permutation equivariant. Since $\rho$ is also pointwise, by a similar reasoning we have $\tilde{\mathbf{y}}_t = \rho(\mathbf{C}(\tilde{\mathbf{S}})\tilde{\mathbf{z}}_t) = \mathbf{P}^{\mathsf{T}}\rho(\mathbf{C}(\mathbf{S})\mathbf{z}_t) = \mathbf{P}^{\mathsf{T}}\mathbf{y}_t$. ■

# References


[1] F. Gama, J. Bruna, and A. Ribeiro, "Stability properties of graph neural networks," *arXiv:1905.04497v2 [cs.LG]*, 4 Sep. 2019. [Online]. Available: http://arxiv.org/abs/1905.04497