The Joint Spectral Radius of Neural Networks

Piero Deidda, Nicola Guglielmi, Francesco Tudisco

Abstract

Several popular architectures of neural networks, e.g. MLPs, graph NNs, RNNs and others can be interpreted as a discrete nonlinear switched system

$$x_{k+1} = f_{\sigma(k)}(x_k) \qquad k = 1, \dots, N \tag{1}$$

where the map $f_{\sigma(k)}$, defining the k-th layer, is chosen from an infinite family of maps $\mathcal{F} := \{f_i\}_{i \in I}$, according to a switching rule σ that is determined by the training process. Then, different phenomena studied in deep learning can be interpreted and studied from the dynamical systems point of view. For example the phenomenon of "oversmoothing" in graph neural networks and transformers can be interpreted in terms of existence and exponential convergence to low rank fixed points of the system. Similarly, the phenomena of "exploding/vanishing gradients" and "vulnerability to adversarial attacks" can be formalized in terms of instabilities of the system.

In this talk we investigate the stability of neural networks modeled as in (1) in the infinite deep regime. In particular, we focus on models where the layer maps $f_{\sigma(k)}$ are subhomogeneous and preserve a cone and its ordering. These Neural Networks have been recently investigated in relation to different stability issues, e.g. cut-off and oversmoothing phenomena [1, 10] as well as existence and uniqness of fixed points in deep equilibrium models [7, 9]. In particular, these models have the advantage of being naturally non expansive with respect to proper metrics defined on the cone, allowing them to be studied in the framework of nonlinear Perron-Frobenius theory [6].

Since the switching rule in neural networks is a-priori unknown, but is chosen by the training process, we study the stability of the switched system in (1) by considering the worst possible case. To this end, since the layer maps are generally nonlinear, we extend the notion of joint spectral radius (**JSR**) of a bounded family of maps, say $\mathcal{F}$, from the linear case [5] to the nonlinear one

$$\mathrm{JSR}(\mathcal{F}) = \limsup_{k \to \infty} \sup_{f \in \Sigma_k(\mathcal{F})} \|f\|^{\frac{1}{k}}$$

where $\|f\| = \sup_{x : \|x\| \le 1} \|f(x)\|$ and $\Sigma_k(\mathcal{F})$ is the set of maps that are composition of k maps in $\mathcal{F}$. Then, we prove the following result showing that the nonlinear JSR can be used to establish both the asymptotic and local stability or instability of the system.

Theorem. *Let $\mathcal{K}$ be a closed cone and $\mathcal{F}$ a family of continuous, subhomogenenous and order preserving maps on a cone $\mathcal{K}$, then:*

- *If $JSR(\mathcal{F}) < 1$, then the system is asymptotically stable. Contrary if $JSR(\mathcal{F}) > 1$, there exists some divergent orbit.*
- *For any $x, y \in \mathrm{Int}(\mathcal{K})$ and $\epsilon > 0$, there exists $C(x, y, \epsilon)$ s.t. for all $f \in \Sigma_k(\mathcal{F})$ and $k \in \mathbb{N}$*

$$\|f(x) - f(y)\| \le C(x, y, \epsilon)(JSR(\mathcal{F}) + \epsilon)^k \|x - y\|$$

This result naturally finds applications in the study of the behavior of the network gradients and thus relates to phenomena like "exploding/vanishing gradients" and "vulnerability to adversarial attacks". As a consequence, we investigate deeper the properties of the nonlinear JSR.

First, inspired by [3], we show that the joint spectral radius of the subhomogeneous family $\mathcal{F}$ can be bounded from above and below by the nonlinear joint spectral radii of two families, $\mathcal{F}_0$ and $\mathcal{F}_\infty$, of continuous **homogeneous** and order preserving maps

$$\mathrm{JSR}(\mathcal{F}_\infty) \leq \mathrm{JSR}(\mathcal{F}) \leq \mathrm{JSR}(\mathcal{F}_0),$$

where the families $\mathcal{F}_\infty$ and $\mathcal{F}_0$ are the collections of the limits, for λ going to infinity and zero respectively, of $f(\lambda x)/\lambda$ with f that varies in $\mathcal{F}$. Thus, focusing on the case of a bounded family of continuous, homogeneous and order preserving maps, and similarly to the linear setting [2, 8], we prove that the JSR admits two different reformulations, the first in terms of monotone prenorms of the maps in $\mathcal{F}$ and the second in terms of the spectral radii of the maps in the semigroup generated by $\mathcal{F}$

$$\mathrm{JSR}(\mathcal{F}) = \inf_{\substack{\Theta \text{ Monotone} \\ \text{Prenorm}}} \sup_{f\in\mathcal{F}} \Theta(f), \qquad \mathrm{JSR}(\mathcal{F}) = \limsup_{k\to\infty} \sup_{f\in\Sigma_k(\mathcal{F})} \rho(f)^{\frac{1}{k}}.$$

In the last expression a monotone prenorm is an absolutely homogeneous, positive definite functional that preserve the ordering induced by the cone. Moreover, the spectral radius $\rho(f)$ of a continuous, homogeneous and order preserving map, say f, is introduced according to the nonlinear Perron-Frobenius theory [6] and corresponds to the maximum eigenvalue of the map f. Finally, we present an algorithm devoted to the computation of the nonlinear JSR. Our algorithm is inspired by the polytope algorithm used in the linear case [4] and iteratively builds a monotone extremal prenorm for the system in terms of the Minkowski functional of a finitely generated subset of the cone.

References

[1] B. Avelin and A. Karlsson. Deep limits and a cut-off phenomenon for neural networks. *JMLR*, 2022.

[2] M. A. Berger and Y. Wang. Bounded semigroups of matrices. *LINEAR ALGEBRA APPL*, 1992.

[3] R. L. Cavalcante, Q. Liao, and S. Stańczak. Connections between spectral properties of asymptotic mappings and solutions to wireless network problems. *TSP*, 2019.

[4] N. Guglielmi and V. Protasov. Exact computation of joint spectral characteristics of linear operators. *FoCM*, 2013.

[5] R. Jungers. *The joint spectral radius: theory and applications.* Springer, 2009.

[6] B. Lemmens and R. Nussbaum. *Nonlinear Perron-Frobenius Theory.* Cambridge Univ. Press, 2012.

[7] T. J. Piotrowski, R. L. G. Cavalcante, and M. Gabor. Fixed points of nonnegative neural networks. *JMLR*, 2024.

[8] G.-C. Rota and G. Strang. A note on the joint spectral radius. *Indag. Math*, 1960.

[9] P. Sittoni and F. Tudisco. Subhomogeneous deep equilibrium models. In *ICML*, 2024.

[10] K. Zhang, P. Deidda, D. Higham, and F. Tudisco. Rethinking oversmoothing in graph neural networks: A rank-based perspective. *arXiv*, 2025.