On the contractivity of neural differential equations

Nicola Guglielmi, Arturo De Marinis, Anton Savostianov, Stefano Sicilia, Francesco Tudisco
nicola.guglielmi@gssi.it
GSSI - Gran Sasso Science Institute, Viale Francesco Crispi 7, L'Aquila, 67100, Italy

Neural ODEs [1] are ordinary differential equations whose vector field is a neural network.

As all neural networks, neural ODEs are vulnerable to adversarial attacks, i.e. imperceptible perturbations, added to the inputs of a neural network, designed in such a way that the output is affected by large perturbations. For this it is important to make neural ODEs contractive (see e.g [2]). In this talk we propose a novel methodology to solve a key eigenvalue optimization problem which arises in the contractivity analysis of neural ODEs. More specifically we look at contractivity properties of a one layer weight-tied neural ODE

$$\dot{x}(t) = \sigma(Ax(t) + b), \qquad t \in [0, T], \tag{1}$$

where $x : [0, T] \to \mathbb{R}^n$ is the feature vector evolution function, $A \in \mathbb{R}^{n,n}$ and $b \in \mathbb{R}^n$ are the parameters, and $\sigma : \mathbb{R} \to \mathbb{R}$ is the activation function, assumed to be smooth and such that $\sigma'(\mathbb{R}) \subset [m, 1]$, with $0 < m \leq 1$. ($\sigma : \mathbb{R} \to \mathbb{R}^+$ denotes an activation function and for a vector $z \in \mathbb{R}^n$, $\sigma(z) \in \mathbb{R}^n$ has to be interpreted entry-wise). To this aim we are led to study the logarithmic norm of a set of products of type DA, where D is a diagonal matrix such that $\operatorname{diag}(D) \in \sigma'(\mathbb{R}^n)$.

We propose a two-level nested methodology to solve this optimization problem and extend it to the general multilayer - and possibly time-dependent - case

$$\dot{x}(t) = \sigma\left(A_k(t) \ldots \sigma\left(A_1(t)x(t) + b_1(t)\right) \ldots + b_k(t)\right).$$

To illustrate our methodology, we propose several numerical examples, including the stabilizing performance on a one-layer neural ODE applied to the classification of the MNIST handwritten digits dataset.

References

[1] R. T. Q. Chen, Y. Rubanova, J. Bettencourt and D. K. Duvenaud, *Neural Ordinary Differential Equations*, Advances in Neural Information Processing Systems **31** (2018).

[2] E. Celledoni, M. J. Ehrhardt, C. Etmann, R. I. Mclachlan, B. Owren, C.-B. Schonlieb, and F. Sherry, *Structure-Preserving Deep Learning*, European J. Appl. Math. **32** (2021), no. 5, 888–936.

[3] N. Guglielmi, A. De Marinis, A. Savostianov and F. Tudisco, *Contractivity of neural ODEs: an eigenvalue optimization problem*, submitted.