Stability of neural ordinary differential equations

Arturo De Marinis[1,*], Nicola Guglielmi[1], Anton Savostianov[2], Stefano Sicilia[4], and Francesco Tudisco[1,3]

[1] *Gran Sasso Science Institute (GSSI), L'Aquila, Italy*

[2] *RWTH Aachen University, Aachen, Germany*

[3] *The University of Edinburgh, Edinburgh, United Kingdom*

[4] *University of Mons, Mons, Belgium*

Corresponding author: `*arturo.demarinis[at]gssi.it`

Abstract

Neural ODEs [1] are ordinary differential equations whose vector field is a neural network. The numerical integration of a neural ODE defines a deep 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 corresponding to the perturbed input is far away from the output corresponding to the original input. Nevertheless, neural ODEs are ordinary differential equations, thus the stability theory of ODEs can be applied to make neural ODEs robust and stable against adversarial attacks.

Our contribution is in this direction [2, 3]. We consider the neural ODE

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

where $u : [0, 1] \to \mathbb{R}^d$ is the feature vector evolution function, $A \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$ are the parameters, and $\sigma : \mathbb{R} \to \mathbb{R}$ is the activation function, assumed to be smooth and such that $\sigma'(\mathbb{R}) \subset [\alpha, 1]$, with $0 < \alpha \leq 1$.

The stability of the neural ODE (1) depends on the Lipschitz constant L of its flow ϕ, i.e. the map that associates to an initial condition $u(0) = u_0 \in \mathbb{R}^d$ the corresponding solution $u(1) \in \mathbb{R}^d$. We have proved in [2] that

$$L \leq e^{\delta}, \quad \text{with} \quad \delta := \max_{D\in\Omega_\alpha} \mu_2(DA),$$

where $\Omega_\alpha = \{D \in \mathbb{R}^{d\times d} \ : \ D \text{ is diagonal and } \alpha \leq D_{ii} \leq 1, \ \forall i = 1, \dots, d\}$ and μ_2 denotes the logarithmic 2-norm of a matrix. If δ is positive and small, then the neural ODE (1) is stable, while if δ is positive and large, then it is unstable.

Since we have experimentally noticed that δ is positive and not small in applications, like image classification, we have proposed in [3] a perturbation method to enforce the neural ODE to be stable. In particular, given $\bar{\delta} < \delta$, we are able to compute the smallest (in Frobenius norm) perturbation matrix $\Delta \in \mathbb{R}^{d\times d}$ such that

$$\max_{D\in\Omega_\alpha} \mu_2(D(A + \Delta)) = \bar{\delta}.$$

Therefore, the flow $\bar{\phi}$ of the perturbed neural ODE

$$\dot{u}(t) = \sigma((A + \Delta)u(t) + b), \qquad t \in [0, 1],$$

has Lipschitz constant $\bar{L} \leq e^{\bar{\delta}}$, and so the perturbed neural ODE is more stable than the neural ODE (1).

To illustrate our methodology, we compare the performance of two neural ODEs for MNIST and FashionMNIST classification against the Fast Gradient Sign Method (FGSM) attack: the former

trained according to the state-of-the-art training strategy, and the latter trained introducing the stability constraint. Our experiments indicate that the latter shows a significant improvement in robustness against the FGSM attack.

References

[1] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural Ordinary Differential Equations. *Advances in Neural Information Processing Systems 31, 2018.*

[2] Nicola Guglielmi, Arturo De Marinis, Anton Savostianov, and Francesco Tudisco. Contractivity of neural ODEs: an eigenvalue optimization problem. *arXiv preprint arXiv:2402.13092, 2024.*

[3] Arturo De Marinis, Nicola Guglielmi, Stefano Sicilia, and Francesco Tudisco. Stability of neural ODEs by a control over the expansivity of their flows. *arXiv preprint arXiv:2501.10740, 2025.*