Post-Doctoral Project - Expressivity of Implicit neural networks for data-driven stability and safety certificates

Giorgio Valmorbida[1] Akin Delibasi[2]

[1]Inria DISCO, L2S, CentraleSupélec, Université Paris-Saclay
[2]Department of Computer Science, University College London

April 28, 2026

Abstract

This project investigates the expressivity of implicit neural network architectures for computing data-driven stability and safety certificates in dynamical systems. While feedforward neural networks have been used to approximate Lyapunov functions, their expressivity can be limited in high-dimensional or highly nonlinear settings. Implicit neural networks, defined through fixed-point equations, offer enhanced representational capacity and natural connections to complementarity-based control strategies. We propose a systematic comparison of feedforward and implicit architectures using both ReLU and softplus activations, with a focus on their ability to represent generalised quadratic Lyapunov functions. A key methodological contribution is the development of curriculum-based training strategies, including activation smoothness annealing (gradually transitioning from softplus to ReLU during training as a continuation method) and progressive system complexity scaling for interconnected systems. The project further explores learning-based control synthesis, where computed Lyapunov certificates guide safe policy learning in a data-driven setting. This collaboration between Inria DISCO and UCL Computer Science combines expertise in implicit models, convex optimisation, and Lyapunov theory (Inria) with training methodology, learning-based control, and interconnected systems (UCL).

1 Project Description

The deployment of Neural Networks (NNs) as control-law components in dynamical systems necessitates rigorous guarantees of safe and reliable operation [12]. Formal certification of key system properties—such as stability, robustness (e.g., input-to-state stability), and performance or safety indices—can be achieved by constraining the certificate to a suitable functional class (e.g., Lyapunov or storage functions) and computing its parameters via convex optimization techniques. Furthermore, in scenarios where system complexity or limited prior knowledge precludes the availability of accurate mathematical models, certification must be derived directly from data. This data-driven paradigm bypasses the traditional system identification step, thereby enabling end-to-end control design.

For large-scale systems, however, certificates with high-dimensional parameterizations become increasingly difficult to compute, while overly restrictive choices (e.g., structured quadratic functions) may lead to conservative or suboptimal results. Recent advances exploit compositional Lyapunov function constructions, where feedforward neural networks are employed to approximate Lyapunov functions directly from data [7]. These approaches often rely on Lyapunov inequalities derived from Zubov's theorem [18]. Similarly, feedforward architectures have been used to compute polyhedral Control Lyapunov Functions (CLFs) using ReLU activation functions, leveraging their piecewise-affine structure [9]. Such certificates can subsequently be used for safety verification or for synthesizing feedback controllers, for instance via gradient-based policies derived from storage functions.

Despite their computational convenience—particularly due to compatibility with standard backpropagation frameworks—feedforward architectures may exhibit limited expressivity in high-dimensional or highly nonlinear settings. In contrast, implicit neural networks [4] offer enhanced representational capacity and improved memory efficiency [10], which may significantly benefit scalability. When ReLU activations are employed, the resulting mappings are piecewise affine and may exhibit discontinuities, or may present solutions that are only locally well-posed (i.e., defined on compact subsets of the state space). Notably, such implicit representations can be reformulated as Linear Complementarity Problem [11], establishing connections with widely used control strategies such as quadratic-programming and linear-programming-based model predictive control (QP-MPC and LP-MPC) [16, 11].

A smooth approximation of the ReLU function is given by the softplus function, parameterized by k:

$$f(x) = \ln(1 + e^{kx}).$$

The approximation quality is governed by the scalar parameter k, with the ReLU recovered in the limit as $k \to \infty$. The approximation of non-differentiable functions by smooth counterparts has been extensively studied in the context of optimization, particularly to analyze limiting behavior and regularization effects [13]. This project proposes to investigate the trade-offs between ReLU and softplus activations in terms of the expressivity [17, 15] of Lyapunov function approximators. In implicit architectures, ReLU activations can be composed to approximate regularized step functions [5, 14], which may induce ill-posed algebraic loops with solutions confined to finite intervals. By contrast, smooth approximations may ensure continuity; however, they do not guarantee Lipschitz boundedness.

From a control synthesis perspective, admissible control inputs are those that ensure a decrease condition on a Control Lyapunov Function. Alternatively, within a data-driven framework, the control policy may be learned by formulating the objective as a reward maximization problem in the context of *learning-based control* (including reinforcement learning [6] and gradient-based policy optimisation). Recent work has demonstrated that neural networks can jointly learn Lyapunov functions and stabilising controllers [3], while safe reinforcement learning methods exploit Lyapunov certificates to guarantee stability during exploration [2].

A central challenge in training neural networks to approximate Lyapunov functions is the design of effective learning strategies. We propose to adopt *curriculum learning* approaches [1], where the training process is structured in stages of increasing difficulty. In particular, we will investigate *activation smoothness annealing*: rather than selecting between softplus and ReLU a priori, the softplus parameter k is gradually increased during training, starting from a smooth optimisation landscape (k small) and progressively approaching the piecewise-affine ReLU limit ($k \to \infty$). This continuation strategy is theoretically grounded in smooth minimisation frameworks [13] and transforms the static expressivity comparison into a dynamic training methodology. Additional curriculum dimensions include progressive scaling of system complexity (from low-dimensional benchmarks to interconnected systems) and gradual expansion of the training region in state space, where each stage defines an invariant set candidate.

A further motivation for the implicit neural network framework is its capacity to represent *feedback-interconnected systems*. The composition rules for implicit models (cascade, parallel, and feedback) enable the analysis of coupled subsystems, where well-posedness of the feedback loop is guaranteed by spectral conditions on the weight matrix [4]. This naturally extends compositional Lyapunov constructions [7] from feedforward coupling to genuine feedback interconnections, addressing a gap in the current literature.

The project is structured around the following research tasks:

1. Define benchmark classes of nonlinear systems with known smooth or piecewise-smooth Lyapunov functions to enable systematic evaluation. For existing methods, a few examples have been put forward in [8]

2. Investigate neural network architectures that induce Lyapunov function representations in the form of generalized quadratic functions.

3. For Lyapunov function approximation, systematically compare—under equal parameter budgets—the expressivity of feedforward and implicit neural network architectures using both ReLU and softplus activations.

4. Based on the approximation error of the computed Lyapunov functions, assess whether convergence guarantees extend to invariant sets rather than

isolated equilibria, and assess dynamical properties impacted by the approximation errors, such as the safety of trajectories.

5. Develop curriculum-based training strategies [1] for Lyapunov neural networks, including activation smoothness annealing (k-curriculum) and progressive system complexity scaling for interconnected systems.
6. Investigate learning-based control synthesis [3, 2] using computed Lyapunov certificates as safety constraints or shaped objectives for policy learning.

In parallel, the following practical aspects will be addressed:

$P1$ Develop improved algorithms for parameter computation in implicit neural networks, with emphasis on numerical stability and convergence.

$P2$ Perform a comparative analysis of the computational complexity and efficiency of feedforward and implicit architectures, considering both ReLU and softplus activation functions.

2 Work Plan

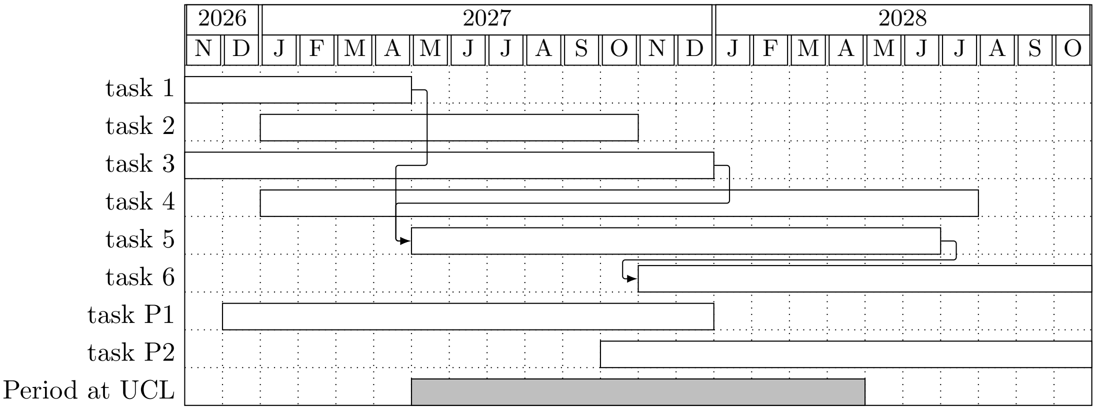

Figure 1: Project schedule.

The expected outcome and deliverables of the project are:

$O1$ Journal papers on the expressivity comparison of feedforward and implicit neural network architectures for Lyapunov function approximation (Tasks 1–3).

- $O2$ Journal papers on curriculum learning strategies for Lyapunov neural network training, including activation smoothness annealing and system complexity scaling (Task 5).
- $O3$ Conference and Journal papers on learning-based control with Lyapunov safety certificates for interconnected systems (Task 6).
- $O4$ Scientific reports on the findings of the two practical tasks P1 and P2.
- $O5$ Open-source software toolbox for computing implicit neural network Lyapunov functions with curriculum-based training.

The post-doc will take place within the Inria Team DISCO, hosted at the L2S Laboratory of Signals and Systems, with a target starting date of November, 1st 2026.

In May 2027, the postdoctoral researcher will start his/her one year stay in London, UK, in collaboration with Prof. Akin Delibasi in the Computer Science department of University College London. The collaboration will focus on the following complementary aspects of the proposed methods:

- **Curriculum-based training methodology (Task 5):** Development of structured training strategies [1] for Lyapunov neural networks. The primary approach is activation smoothness annealing, where the softplus parameter k is gradually increased during training as a continuation method [13]. Supporting strategies include progressive system complexity curricula (scaling from single subsystems to interconnected systems) and state-space expansion curricula (training outward from the equilibrium to define increasingly large invariant sets). This work builds on Delibasi's expertise in AI and machine learning methodology.
- **Interconnected and feedback systems:** Extension of the implicit neural network framework to feedback-connected systems, leveraging the composition rules of implicit models and compositional Lyapunov constructions. The progressive interconnection curriculum provides a principled pathway from isolated subsystems to fully coupled networks. This draws on Delibasi's background in stability analysis of interconnected, LPV, and switched delay systems, and in LMI/PMI-based methods developed during his research at LAAS-CNRS with D. Henrion and D. Arzelier.
- **Learning-based control with safety certificates (Task 6):** Investigation of how computed Lyapunov certificates can guide safe policy learning in a data-driven setting [2], including Lyapunov-constrained policy optimisation and curriculum-guided safe exploration.

The two-site structure reflects the logical organisation of the research: the Inria phase (Tasks 1–4, P1–P2) develops the benchmark systems, implicit neural network architectures, expressivity theory, and the Lyapunov analysis for stability and safety, while the UCL phase (Tasks 5–6) develops the training methodology, interconnected system extensions, and learning-based control synthesis. Both phases are essential to the project: without the Inria contributions, there are no architectures or expressivity results to build upon; without the UCL contributions, the training strategies remain standard, and the Lyapunov certificates lack a pathway to data-driven controller design.

Collaboration context

Both Valmorbida and Delibasi held research positions at LAAS-CNRS, Toulouse, sharing a common research heritage in polynomial optimisation, Lyapunov methods, and convex optimisation for control. This postdoctoral project would be among the first funded positions under the UCL-Inria Joint Centre agreement signed in 2025, demonstrating that the partnership is producing concrete collaborative research in artificial intelligence and control.

References

[1] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th International Conference on Machine Learning (ICML)*, pages 41–48. ACM, 2009.

[2] Felix Berkenkamp, Matteo Turchetta, Angela P. Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017.

[3] Ya-Chien Chang, Nima Roohi, and Sicun Gao. Neural lyapunov control. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

[4] Laurent El Ghaoui, Fangda Gu, Bertrand Travacca, Armin Askari, and Alicia Tsai. Implicit deep learning. *SIAM Journal on Mathematics of Data Science*, 3(3):930–958, 2021.

[5] Francesco Ferrante and Giorgio Valmorbida. Stability analysis of a class of discontinuous discrete-time systems. *IEEE Control Systems Letters*, 7:454–459, 2023.

[6] Javier García and Fernando Fernández. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(42):1437–1480, 2015.

[7] Lars Grüne. Computing lyapunov functions using deep neural networks. *Journal of Computational Dynamics*, 8(2):131–152, 2021.

[8] Lars Grüne and Mario Sperl. Examples for separable control lyapunov functions and their neural network approximation. *IFAC-PapersOnLine*,

56(1):19–24, 2023. 12th IFAC Symposium on Nonlinear Control Systems NOLCOS 2022.

[9] Lars Grüne, Mario Sperl, and Debasish Chatterjee. Representation of practical nonsmooth control lyapunov functions by piecewise affine functions and neural networks. *Systems Control Letters*, 202:106103, 2025.

[10] Ingo Gühring, Mones Raslan, and Gitta Kutyniok. *Expressivity of Deep Neural Networks*, page 149–199. Cambridge University Press, 2022.

[11] Morten Hovd and Giorgio Valmorbida. Solving LP-MPC problems using ramp functions. In *2023 IEEE Conference on Control Technology and Applications (CCTA)*, pages 445–450, 2023.

[12] Ian R. Manchester, Ruigang Wang, and Nicholas H. Barbara. Neural networks in the loop: Learning with stability and robustness guarantees. *Annual Review of Control, Robotics, and Autonomous Systems*, 2026.

[13] Yu. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

[14] Armin Nurkanović, Anton Pozharskiy, Jonathan Frey, and Moritz Diehl. Finite elements with switch detection for numerical optimal control of nonsmooth dynamical systems with set-valued heaviside step functions. *Nonlinear Analysis: Hybrid Systems*, 54:101518, 2024.

[15] Philipp Petersen and Felix Voigtlaender. Optimal approximation of piecewise smooth functions using deep relu neural networks. *Neural Networks*, 108:296–330, 2018.

[16] Giorgio Valmorbida and Morten Hovd. Quadratic programming with ramp functions and fast online QP-MPC solutions. *Automatica*, 153:111011, 2023.

[17] Dmitry Yarotsky. Error bounds for approximations with deep relu networks. *Neural Networks*, 94:103–114, 2017.

[18] Vladimir Ivanovich Zubov et al. *Methods of AM Lyapunov and their application*. P. Noordhoff LTD, Groningen, The Netherlands (1964), 1964.