Identifiability and guaranteed recovery of polynomial and tensor networks

Master's internship proposal

Context

Modern language and vision models, such as transformers [1] and mixture-of-experts (MoE) architectures [2], have billions of parameters, and their performance improves with model and training data size. However, the neural network parameters are far from unique: permutations of neurons, heads or experts, rescalings and rotations of hidden representations leave the computed function unchanged. Accounting for these parameter symmetries can be beneficial for improving the performance of the models [3], their compression and training [4], optimization [5] and interpretability [6, 7]. Model merging [8, 3] is a typical example: several models trained or fine-tuned independently are combined into one by averaging or adding their weight updates, without retraining; aligning model symmetries is therefore crucial. These topics feature in recent workshops such as UniReps (NeurIPS 2025) and U&ME (ECCV 2026). One of the fundamental questions is identifiability: *when are the weights determined by the input–output map up to symmetries, and can they be recovered efficiently?*

Tensor decompositions and, more generally, tensor networks are a powerful approach to representing higher-order arrays and functions, with rich geometry and approximation theory [9, 10, 11, 12], and they attract increasing interest in the machine learning community (see tutorials at AAAI 2025 and NeurIPS 2025, and the CoLorAI workshop (ICML 2026)). Besides their use in compression, they also play an important role in studying theoretical properties of deep representations [13]: the classic work of Cohen and Shashua [14] studies expressivity, while recent work showed that identifiability can be studied using tensor methods [15], for a special case of polynomial activations.

The internship further explores this connection between tensor networks and polynomial neural networks, with the goal of studying identifiability and symmetries of models in realistic settings. We aim at exploiting recent theoretical advances [16] in the context of larger models, with a view towards compression, model merging, and guaranteed weight recovery.

Candidate profile

Master 2 or final-year engineering student in applied mathematics, signal processing, machine learning or a related field. A strong background in linear algebra and optimization and programming skills in Python are expected; prior knowledge of tensor methods or deep learning theory is a plus but not required. Good written and spoken English.

Work conditions

The internship will take place in the Simul research group of the Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, in Nancy, France (https://cran-simul.github.io). It is part of the TENSOR4ML project of the PEPR IA programme, and the intern will interact with project members. It will be supervised by Konstantin Usevich, Ricardo Borsoi and Marianne Clausel. Duration: about six months, starting early 2027 (dates to be confirmed). The intern receives the standard internship allowance.

Please send your applications (CV, motivation letter and references) to konstantin.usevich@univ-lorraine.fr, ricardo.borsoi@univ-lorraine.fr and marianne.clausel@univ-lorraine.fr . The internship may be followed by a PhD in the team, subject to funding.

References

[1] A. Vaswani et al., Attention is all you need, *NeurIPS*, 2017.

[2] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, J. Dean, Outrageously large neural networks: the sparsely-gated mixture-of-experts layer, *ICLR*, 2017.

[3] A. Theus, A. Cabodi, S. Anagnostidis, A. Orvieto, S. P. Singh, V. Boeva, Generalized linear mode connectivity for transformers, *NeurIPS*, 2025 (oral).

[4] Z. Liu et al., SpinQuant: LLM quantization with learned rotations, *ICLR*, 2025.

[5] B. Zhao, N. Dehmamy, R. Walters, R. Yu, Symmetry teleportation for accelerated optimization, *NeurIPS*, 2022.

[6] M. Pearce, T. Dooms, A. Rigg, J. Oramas, L. Sharkey, Bilinear MLPs enable weight-based mechanistic interpretability, *ICLR*, 2025.

[7] W. Nelson, T. Karaletsos, F. Locatello, Toward identifiable sparse autoencoders, *ICML*, 2026.

[8] V. Bürgin, D. Herbst, Y.-W. E. Lin, S. Jegelka, Beyond structural symmetries: linear mode connectivity via neuron identifiability, *ICML*, 2026.

[9] J. M. Landsberg, *Tensors: Geometry and Applications*, Graduate Studies in Mathematics 128, AMS, 2012.

[10] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, Springer Series in Computational Mathematics 42, Springer, 2012.

[11] A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, D. P. Mandic, Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions, *Foundations and Trends in Machine Learning*, 9(4–5):249–429, 2016.

[12] A. Cichocki, A.-H. Phan, Q. Zhao, N. Lee, I. Oseledets, M. Sugiyama, D. P. Mandic, Tensor networks for dimensionality reduction and large-scale optimization: Part 2 applications and future perspectives, *Foundations and Trends in Machine Learning*, 9(6):431–673, 2017.

[13] R. Borsoi, K. Usevich, M. Clausel, Low-rank tensor decompositions for the theory of neural networks, *IEEE Signal Processing Magazine*, arXiv:2508.18408, 2026.

[14] N. Cohen, O. Sharir, A. Shashua, On the expressive power of deep learning: a tensor analysis, *COLT*, 2016.

[15] K. Usevich, R. Borsoi, C. Dérand, M. Clausel, Identifiability of deep polynomial neural networks, *NeurIPS*, 2025 (oral), arXiv:2506.17093.

[16] G. L. Marchetti, V. Shahverdi, S. Mereta, M. Trager, K. Kohn, Position: algebra unveils deep learning – an invitation to neuroalgebraic geometry, *ICML*, 2025.