Sparse polynomial optimization with applications to deep networks

Jean-Bernard Lasserre (LAAS-CNRS)
Victor Magron (LAAS-CNRS)
Jun Zhao (NTU, Singapore)

September 28, 2021

Context: Modern state of the art predictive neural architectures are known to be vulnerable to small perturbation of the input data [2]. This is an important issue for critical applications. First attempt to circumvent this problems are based on the adaptation of robust optimization techniques to deep network architectures. These approaches are mostly empirical and it has been demonstrated that many such adversarial training strategies can be circumvented [1]. An alternative to empirical approaches is to certify robustness against adversarial attacks. We will be mostly interested in convex relaxation based certification approaches such as LP relaxations [3] and SDP relaxations [5]. A tutorial on state of the art in this field is given in [4].

Goal of the PhD thesis: Lasserre's Hierarchy is a generic tool which can be used to solve global polynomial optimization problems over semi-algebraic sets [11]. It has many application among which providing numerical certificates for formal proofs and certification [8]. In a first step we would like to investigate how this approach could be used to extend the approaches of [5, 6] to provide a hierarchy of adversarial robustness bounds for convolutional deep networks. In a second step, we will investigate how to go beyond previous works [7] to perform more accurate stability analysis of recurrent neural networks. We will exploit the specific structure of neural network robustness certification problems, such as sparsity of symmetries, to devise efficient numerical algorithms to scale the approach up to real world networks. From an academic point of view, the fruitful synergy between the existing methodology of relaxation hierarchies and the scientific challenge of robustness certification is expected to impact at the methodological level with development of novel tools and the application level with the conception of new practical solutions.

Potential industrial impact: Recent collaborations [9, 10] between INRIA, CNRS and RTE allowed to successfully apply sparse variants of Lasserre's hierarchies. In the context of energy networks, the hierarchies are currently used to solve industrial-scale power flow problems with thousands of variables. Such Industrial application emerged successfully because of appropriate interplay between methodology and practice. This situation shares a lot of similarities with the elements described in the present project and we expect similar breakthroughs. Certification of systems involving data driven trained components is a middle term goal in the transport industry. We propose to develop a method enabling certification of robustness margins of AI components of such systems, when considered as black box input/output systems. Developing such tools is a necessary first step toward the certification of systems involving AI trained components and a large degree of autonomy. Such systems are expected to be more and more common in the transport industry and constitute a major challenge in terms of certification.

Requirements: A successful candidate will have a strong background in applied mathematics or computer science, having a very good knowledge of probability and statistics as well as a working knowledge of convex optimization, real analysis and basic measure theory. The candidate is expected to have strong programming skills, be highly motivated and creative.

Funding: This PhD will be funded by DesCartes (A CREATE Programme on AI-based Decision making in Critical Urban Systems), a hybrid AI project between CNRS and Singapore. It will be co-supervised between Nanyang Technological University (NTU), Singapore and LAAS CNRS. The PhD candidate will be hosted in NTU, Singapore.

References

[1] A. Athalye, N. Carlini and D. Wagner (2018). Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. ICML.

[2] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus (2014). Intriguing properties of neural networks. In International Conference on Learning Representations

[3] E. Wong and J. Z. Kolter (2018). Provable defenses against adversarial examples via the convex outer adversarial polytope. ICML.

[4] Z Kolter, Madry (2018). Adversarial Robustness: Theory and Practice. NIPS tutorial, `adversarial-ml-tutorial.org`

[5] T. Chen, J.-B. Lasserre, V. Magron, and E. Pauwels (2020). *Semialgebraic Optimization for Bounding Lipschitz Constants of ReLU Networks*. NeurIPS.

[6] T. Chen, J.-B. Lasserre, V. Magron, and E. Pauwels (2021). *Semialgebraic Representation of Monotone Deep Equilibrium Models and Applications to Certification*. arXiv:2106.01453.

[7] Y. Ebihara, H. Waki, V. Magron, N. H. A. Mai, D. Peaucelle, and S. Tarbouriech (2021). *Stability Analysis of Recurrent Neural Networks by IQC with Copositive Mutipliers*. Proceedings of the Control and Decision Conference.

[8] V. magron, X. Allamigeon, S.Gaubert and B.Werner (2015). Formal Proofs for Nonlinear Optimization. Journal of Formalized Reasoning Vol, 8(1), 1-24.

[9] C. Josz. *Application of polynomial optimization to electricity transmission networks*. Theses, Université Pierre et Marie Curie - Paris VI, July 2016.

[10] J. Wang and V. Magron (2021). *Certifying Global Optimality of AC-OPF Solutions via the CS-TSSOS Hierarchy*. arXiv:2109.10005.

[11] J.B. Lasserre (2001). Global optimization with polynomials and the problem of moments. SIAM Journal on optimization, 11(3), 796-817.