



Institut de Recherche
en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J



Institut de Recherche en Informatique de Toulouse

Équipe Signaux & Communications
<http://sc.enseeiht.fr/>

Contact : Henrique Goulart
henrique.goulart@irit.fr

Master 2 internship proposal – Spring 2021 (5 to 6 months) : Estimation of large-dimensional tensor models and applications in machine learning

Supervisors

The internship will be supervised by Prof. H. Goulart (henrique.goulart@irit.fr) at IRIT/ENSEEIHT, in remote collaboration with two members of the LargeDATA (DataScience) chair at 3IA MIAI/Univ. Grenoble-Alpes : Prof. R. Couillet (head, romain.couillet@gipsa-lab.grenoble-inp.fr) and Dr. P. Comon (pierre.comon@gipsa-lab.grenoble-inp.fr).

Funding

This internship will be funded by the Artificial and Natural Intelligence Toulouse Institute (3IA ANITI, <http://aniti.univ-toulouse.fr>), as part of the AI Research Chair lead by N. Dobiéon.

Context

Tensor models are powerful tools for addressing many problems in signal processing, machine learning and beyond [1]–[3]. Yet, their use in these applications typically requires estimating a low-rank tensor from a set of observations corrupted by noise, which is often a difficult task. Moreover, in most cases there is currently no theory for predicting the actual estimation performance that can be attained.

To overcome this gap, in recent years several researchers have studied the asymptotic statistical performance of ideal and practical estimators in the large-dimensional regime, where the size of the tensor grows large [4]–[6]. In particular, these works have uncovered the abrupt phase transition that the performance of an ideal estimator may undergo as the signal-to-noise ratio grows (see Figure 1 below for an illustration). While some important advancements have been achieved, many scenarios of practical interest remain unexplored, as well as the practical implications of the existing results in applications.

Objectives

The overall goal of this internship is to study extensions and applications of the existing results, as a first step for pushing the existing theory beyond its current limits. We will in particular consider extensions to more general tensor models that apply to larger classes of real-world problems, including e.g. asymmetric models. Application to practical machine learning problems—such as community detection in hypergraphs [7], latent variable model estimation [2] and high-order co-clustering [8]—will also be considered.

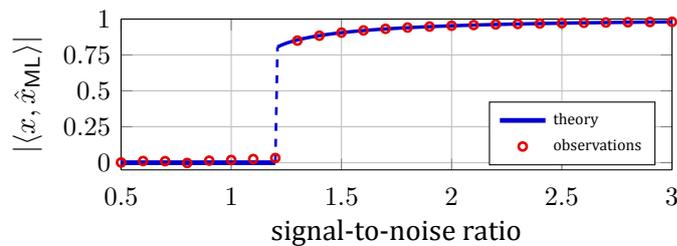


Figure 1: Asymptotic alignment between the maximum likelihood estimate \hat{x}_{ML} and the planted vector x of the symmetric rank-one third-order spiked tensor model, as predicted by [4].

The intern will initially perform computer simulations aimed at understanding the behavior of ideal and practical estimators in the target scenarios/applications. Some theoretical results may then be derived on the basis of these experimental findings. Scientific dissemination of these findings will also be encouraged, via publication of papers and/or participation in scientific events.

A PhD thesis may be proposed to the intern at the end.

Candidate profile

We look for strongly motivated candidates with a solid background on mathematics and statistics, having good programming skills in scientific computing languages (Python, Matlab, Julia). Basic knowledge or interest in random matrix theory is a strong plus.

Practical information

- The intern will be hosted at the ENSEEIHT site of IRIT, located at 2 rue Charles Camichel, Toulouse.
- The monthly internship gratification is of 600€.
- Application procedure: please send your CV, your report card and a motivation letter to Henrique Goulart (henrique.goulart@irit.fr).

References

- [1] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, *et al.*, “Tensor decomposition for signal processing and machine learning,” *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.
- [2] A. Anandkumar, D. Hsu, S. M. Kakade, *et al.*, “Tensor decompositions for learning latent variable models,” *Journal of Machine Learning Research*, vol. 15, pp. 2773–2832, 2014.
- [3] S. Rabanser, O. Shchur, and S. Günnemann, “Introduction to tensor decompositions and their applications in machine learning,” *arXiv preprint arXiv:1711.10781*, 2017.
- [4] A. Jagannath, P. Lopatto, and L. Miolane, “Statistical thresholds for Tensor PCA,” *The Annals of Applied Probability*, vol. 30, no. 4, pp. 1910–1933, 2020.
- [5] A. Perry, A. S. Wein, and A. S. Bandeira, “Statistical limits of spiked tensor models,” *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, vol. 56, no. 1, pp. 230–264, 2020.
- [6] A. Montanari and E. Richard, “A statistical model for tensor PCA,” *Advances in Neural Information Processing Systems*, vol. 4, pp. 2897–2905, 2014. eprint: 1411.1076.
- [7] C. Kim, A. S. Bandeira, and M. X. Goemans, “Community detection in hypergraphs, spiked tensor models, and sum-of-squares,” in *2017 International Conference on Sampling Theory and Applications (SampTA)*, Tallinn, Estonia, Jul. 2017, pp. 124–128.
- [8] E. E. Papalexakis, N. D. Sidiropoulos, and R. Bro, “From k-means to higher-way co-clustering: Multilinear decomposition with sparse latent factors,” *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 493–506, 2013.