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

INSTITUT de MATHEMATIQUES de TOULOUSE

M2 Internship + PhD opportunity:

Understanding and improving low-rank tensor model estimators in modern machine learning regimes

Keywords. Random tensor models, machine learning theory, high-dimensional statistics, random matrix theory.

Duration. 5 to 6 months, starting between February and May 2026.

Funding. This internship will be fully funded by the ANR JCJC project "LATENT" led by Henrique Goulart (`https://www.irit.fr/~Henrique.Goulart/latent/`). A 3-year PhD opportunity on the same subject is available at the end of the internship.

Location. Institut de Recherche en Informatique de Toulouse (IRIT), ENSEEIHT site (2 rue Charles Camichel, Toulouse).

Context. Numerous problems in various domains, notably in data science, machine learning and signal processing, can be addressed by resorting to a low-rank tensor model [13, 2, 12, 8, 3]. In general, this requires estimating the parameters of such a model from a given noisy data tensor, which is assumed to contain a low-rank signal bearing the information of interest. Even though several existing algorithms are often capable of accomplishing this notoriously difficult task in a satisfying manner, it is hard to anticipate or guarantee their actual performance in practice.

Significant progress has been achieved in recent years by considering this estimation problem in the large-dimensional regime where the tensor dimensions grow large, which is prevalent in modern applications. Notably, this approach has given access to the exact asymptotic performance of several algorithms aimed at estimating a rank-one signal planted in a noisy tensor [6, 7, 11, 5]. In particular, one of the supervisors of this internship has recently developed an approach based on tools from random matrix theory (RMT) for the study of random tensor models and their estimation [6]. This contribution is related to a growing number of works exploiting the powerful framework of RMT to better understand and improve large-data learning [1, 4, 9, 10]. However, to this day existing results are mostly confined to the rank-one setting, with a few exceptions that do not cover many tensor models and algorithms of practical interest.

This M2 internship is situated at the interface between mathematics and computer science. More specifically, it will take place in the context of project LATENT, which has two ambitious goals: (i) mathematically analyzing the performance and limitations of several tensor model estimation algorithms of practical relevance, beyond the rank-one setting; (ii) proposing novel, improved algorithms with performance guarantees.

Objective and expected work. This M2 internship will focus on the estimation of a low-multilinear-rank tensor model, also known as a low-rank Tucker decomposition, which is a widespread tool for dimensionality reduction of tensor data. In particular, it is very often used for reducing the size of a canonical polyadic decomposition problem [13].

Specifically, we will aim both at studying the performance of existing algorithms in the large-dimensional regime, and at proposing new, improved algorithms. The mathematical analysis of the algorithms will largely rely on tools from RMT and from high-dimensional probability and statistics.

Accomplishing this objective will require both computational work, consisting in performing numerical simulations and interpreting their results, and theoretical work, in the form of a mathematical analysis of the algorithms of interest and their related models.

Internship supervision and scientific environment. The internship will be co-supervised by Henrique Goulart (Toulouse INP, IRIT) and Xiaoyi Mai (Université de Toulouse Jean Jaurès, IMT). Together they gather expertise in tensor & matrix decompositions, high-dimensional statistics, signal processing and machine learning.

The student will be mainly hosted by the Signal and Communications team of IRIT, but part of the work may be carried out at the Statistics and Optimization team of IMT. They will therefore benefit from a scientifically rich environment, with opportunities for exchange and interaction with members of two research institutions having strong expertise in machine learning, high-dimensional statistics and signal processing.

Candidate profile. We look for strongly motivated candidates:

- coming from a math, physics, computer science or engineering diploma;
- having a strong mathematical background, notably in linear algebra, analysis, probability and statistics;
- having good programming skills on some scientific language, preferably python, julia or Matlab.

Knowledge of random matrix theory and of tensor (multilinear) algebra is not required, but is a strong plus. Knowledge of basic optimization theory is also appreciated.

Practical information.

- The intern will be mainly hosted at the ENSEEIHT site of IRIT (2 rue Charles Camichel, Toulouse), in a lively neighborhood close to the city center. However, he/she may also spend some periods at the Université Paul Sabatier site of IMT, located at 118 route de Narbonne, Toulouse.
- The monthly internship gratification is of about 670€.
- Application procedure: send a motivation letter, a CV and your University transcript (relevé de notes) to henrique.goulart@irit.fr and xiaoyi.mai@math.univ-toulouse.fr.

References

[1] M. S. Advani, A. M. Saxe, and H. Sompolinsky. High-dimensional dynamics of generalization error in neural networks. *Neural Networks*, 132:428–446, 2020.

[2] A. Anandkumar, D. Hsu, S. M. Kakade, and M. Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15:2773–2832, 2014.

[3] X. Bi, X. Tang, Y. Yuan, Y. Zhang, and A. Qu. Tensors in statistics. *Annual review of statistics and its application*, 8(1):345–368, 2021.

[4] R. Couillet and Z. Liao. *Random Matrix Methods for Machine Learning*. Cambridge University Press, 2022.

[5] M. Feldman. and D. Donoho. Sharp recovery thresholds of tensor PCA spectral algorithms. In *NeurIPS*, Vancouver, Canada, 2024.

[6] J. H. de M. Goulart, R. Couillet, and P. Comon. A random matrix perspective on random tensors. *Journal of Machine Learning Research*, 23:1–36, 2022.

[7] A. Jagannath, P. Lopatto, and L. Miolane. Statistical thresholds for Tensor PCA. *The Annals of Applied Probability*, 30(4):1910–1933, 2020.

[8] M. Janzamin, H. Sedghi, and A. Anandkumar. Beating the perils of non-convexity: Guaranteed training of neural networks using tensor methods. *arXiv preprint arXiv:1506.08473*, 2015.

[9] X. Mai and R. Couillet. A random matrix analysis and improvement of semi-supervised learning for large dimensional data. *Journal of Machine Learning Research*, 19(1):3074–3100, 2018.

[10] X. Mai and R. Couillet. Consistent semi-supervised graph regularization for high dimensional data. *Journal of Machine Learning Research*, 22:94–1, 2021.

[11] A. Montanari and E. Richard. A statistical model for tensor PCA. *Advances in Neural Information Processing Systems*, 4:2897–2905, 2014.

[12] S. Rabanser, O. Shchur, and S. Günnemann. Introduction to tensor decompositions and their applications in machine learning. *arXiv preprint arXiv:1711.10781*, 2017.

[13] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017.