

# **Workshop: Recent developments beyond classical regimes in statistical learning**

**Monday, November 4, 2024 - Friday, November 8, 2024**

**Institut de Mathématiques de Toulouse (IMT)**

## **Scientific Program**

**Jean Barbier (ICTP, Italy) - Denoising-factorisation phase transition in extensive rank symmetric matrix factorisation**

**Abstract:** Matrix factorisation is central to signal processing and machine learning. Despite many attempts, its statistical analysis in the highly relevant regime where the matrix to infer has a rank growing proportionally to its dimension has remained a challenge, except when the signal is rotationally-invariant. Beyond this setting few results can be found. The reason is that the problem is not a usual spin system because of the growing rank dimension, nor a matrix model (as appearing in high-energy) due to the lack of rotation symmetry, but rather an hybrid between the two.

I will present recent progress towards the understanding of matrix factorisation in a Bayesian setting which does not assume rotational invariance. Using Monte Carlo simulations we draw conclusions about the phase diagram. These pinpoint a denoising-factorisation transition separating a phase where factorisation is not possible but denoising is and universality properties hold, of the same nature as in random matrix theory, from one where factorisation is possible but algorithmically hard, and universality breaks down. We then combine mean-field techniques in an interpretable multiscale fashion in order to access the minimum mean-square error and mutual information. The theory matches well the numerics when accounting for finite size effects.

**Giulio Biroli (ENS, France) - Generative AI and Diffusion Models - A Statistical Physics Analysis**

**Charles Bordenave (IMM, France) - Freeness for tensors**

**Abstract:** In this joint work with Rémi Bonnin, we will lay the foundations of a free probability theory for tensors and establish its relevance in the study of random tensors of high dimension. We will give a definition of freeness associated to a collection of tensors of possibly different orders. We will present the combinatorial theory of free cumulants which are associated to this notion of tensor freeness. Finally, we will see that the basic models of random tensors are asymptotically free as the dimension goes to infinity.

**Aurélien Decelle (UCM, Spain) - How phase transitions shape the learning of complex data in the Restricted Boltzmann Machine**

**Franck Lutzeler (IMT, France) - What is the long-run distribution of stochastic gradient descent? A large deviations analysis**

**Abstract:** We examine the long-run distribution of stochastic gradient descent (SGD) in general, non-convex problems. Specifically, we seek to understand which regions of the problem's state space are more likely to be visited by SGD, and by how much. Using an approach based on the theory of large deviations and randomly perturbed dynamical systems, we show that the long-run distribution of SGD resembles the Boltzmann-Gibbs distribution of equilibrium thermodynamics with temperature equal to the method's step-size and energy levels determined by the problem's objective and the statistics of the noise.

Joint work w/ W. Azizian, J. Malick, P. Mertikopoulos

**Aukosh Jagannath (Waterloo U., Canada) - Effective dynamics and spectral alignment**

**Jon Keating (Oxford U., UK) - Some connections between random matrix theory and machine learning**

**Abstract:** I will discuss some connections between random matrix theory and machine learning, focusing on the spectrum of the hessian of the loss surface.

**Bertrand Lacroix-A-Chez-Toine (KCL, UK) - Random landscape built by superposition of plan waves in high dimension**

**Marc Lelarge (ENS, France) - Combinatorial Optimization with Graph Neural Network: chaining to learn the Graph Alignment Problem**

**Cosme Louart ( Hong Kong U., China) - Operation with concentration inequalities in high dimension**

**Abstract:** In this talk we will present new results to trace concentration inequalities through Lipschitz but also non-Lipschitz functionals. The flexibility of our approach shows that the same mechanism allows to treat similarly concentrated vectors whose observation tails have exponential decay, up to those which do not admit finite moments. We will give some precise and natural examples of such heavy-tailed vectors in high dimension. We will then illustrate our results with selected applications in random matrix theory and machine learning.

**Bruno Loureiro (ENS, France) - Learning features with two-layers neural networks, one step at time**

**Nicolas Macris (EPFL, Switzerland) - Sampling diffusion process**

**Pascal Maillard (IMT, France) - Probing the transition from polynomial to exponential complexity in spin glasses via N-particle branching Brownian motions**

**Subhabrata Sen (Harvard U., USA) - Causal effect estimation under inference using mean field methods**

**Abstract:** We will discuss causal effect estimation from observational data under interference. We adopt the chain-graph formalism of Tchetgen-Tchetgen et. al. (2021). Under “mean-field” assumptions on the interaction networks, we will introduce novel algorithms for causal effect estimation using Naive Mean Field approximations and Approximate Message Passing. Our algorithms are provably consistent under a “high-temperature” assumption on the underlying model. Finally, we will discuss parameter estimation in these models using maximum pseudo-likelihood, and establish the consistency of the downstream plug-in estimator.

Based on joint work with Sohom Bhattacharya (U Florida).

**Inbar Seroussi (Tel Aviv U., Israel) - Exact Dynamics of Stochastic and Adaptive Optimization in High Dimension with Structured Data**

**Ludovic Stephan (ENSAI, France) - A non-backtracking method for long matrix and tensor completion**

**Christos Thrampoulidis (British Columbia U., Canada) - On the Implicit Geometry of Word and Context Embeddings in Next-token Prediction**

**Abstract:** The talk explores optimization principles of next-token prediction (NTP), which has become the go-to paradigm for training modern language models. We frame NTP as cross-entropy optimization across distinct contexts, each tied to a sparse conditional probability distribution across a finite vocabulary. This leads us to introduce "NTP-separability conditions," which enable reaching the entropy lower bound of the NTP objective. We then focus on NTP-trained linear models for which we fully specify the optimization bias of gradient descent. Our analysis highlights the key role played by the sparsity pattern of the contexts' conditional distributions and introduces a NTP-specific notion of margin. We also investigate a log-bilinear NTP model, which abstracts sufficiently expressive language models: In large embedding spaces, we can characterize the geometry of word and context embeddings in relation to a NTP-margin-maximizing logit matrix, which separates in-support from out-of-support words. Through experiments we show how this optimization perspective establishes new links between geometric properties of the embeddings and textural structures as encoded in the sparsity patterns of language.

**Malik Tiomoko (Huawei France) - Enhancing Time Series Forecasting with Random Matrix Theory**

**Abstract:** This talk delves into the application of Random Matrix Theory (RMT) to enhance time series forecasting models. The presentation is structured into two main parts.

In the first part, we analyze the Echo state Neural Network(ESN), a popular time series analysis algorithm using RMT to identify critical data statistics and hyperparameters that influence its performance. By leveraging RMT to theoretically understand the model's dynamics and optimizing its hyperparameters, we aim to significantly improve ESN's forecasting capabilities.

In the second part, we demonstrate how RMT can extend any univariate forecasting model to handle multivariate time series. We frame the multivariate time series forecasting problem as a multi task learning problem, analyze it theoretically in a simplified case, and derive key insights. These insights lead to practical improvements that enable univariate models to effectively manage multivariate data.

**Pierfrancesco Urbani (IPHT, France) - Statistical physics of learning in high-dimensional chaotic systems**

**Abstract:** Recurrent neural networks can be regarded as simple models of the building blocks of microcircuits in the brain. When the synaptic connections between neurons are extracted at random, these models exhibit chaotic dynamical phases. How to train such systems to perform

given tasks is not clear and recently some algorithms have been proposed.

In this talk I will describe this problem and adapt it to simplified high-dimensional non-linear chaotic systems. I will then show that one can study a set of learning algorithms in the large dimensional limit via dynamical mean field theory. This allows to control the statistical properties of the dynamical attractors where the dynamics lands. If time permits, I will also discuss how one can use chaotic noise as a source of statistical variability that can be employed for generative tasks.

Short talks:

Jad Hamdan (University of Oxford, UK): Graph expansion of deep neural networks and their universal scaling limits

Mustapha Maimouni (Université Mohamed V, Morocco): Case study - A hybrid approach inspired by artificial neural networks for RFID network planning

Duranthon (EPFL, Switzerland): Generalization in single-layer graph convolutional network