Random matrix theory (RMT) is a branch of mathematics that studies spectral properties of large matrices belonging to different symmetry classes and having entries drawn from different probability distributions. In addition to probability theory, algebra and analysis, it connects to other mathematical fields, such as number theory, combinatorics, asymptotic geometric analysis, statistics, and random graph theory. It has many important applications in physics, statistics, engineering, computer science and beyond and helps to model the behavior of complex systems e.g., nuclear spectra, quantum chaos, wireless communication, data sets, etc. One of the secrets of its popularity is that RMT can capture universal features of complex systems that are independent of the specific details of the model and reveal new insights into their behavior.

Recently, RMT has found its application in the rapidly developing area of Machine Learning and Deep Neural Networks (DNNs), which have significant implications for data analysis. DNNs are artificial neural networks that have multiple layers of neurons between the input and output layers. Being trained using gradient-based optimization methods, they can learn complex patterns and features from large amounts of data, and achieve state-of-the-art results in tasks such as computer vision, natural language processing, speech recognition, and machine translation. But training and understanding DNNs are very challenging because they require large datasets, powerful hardware, careful hyperparameter tuning, and effective regularization techniques. Also they are still poorly understood from a theoretical point of view, which typically makes the implementation of DNNs more art than science. Thus, there is a pressing need for an in-depth mathematical understanding of the behavior of DNNs necessary for the improvement of existing algorithms and the development of new ones.

The analysis of DNNs depends on the enormous number of parameters contained within large random matrices, such as input-output Jacobian, the data covariance matrix of deep neural networks, Hessian and the Fisher Information Matrix. This makes RMT well-suited for the development of novel theoretical foundations for improving the performance of DNNs. Conversely, the problems arising in Deep Learning pose new fundamental challenges for RMT leading to fruitful and promising synergy between these two fields.

Using RMT approaches, a number of important results were obtained regarding the influence of incoming DNN information on outgoing information. It has been shown that initializing DNN matrices with Gaussian random weights guarantees the absence of both exponential growth and exponential decay of gradients. This, in turn, allows in many cases to increase the speed of learning process by some orders of magnitude. RMT can also help to obtain the DNN learning criterion. Numerous studies have shown that over-training a network negatively affects its properties, as does under-training. There is a reason to believe that analyzing the singular values of the network Jacobian will make it possible to develop optimal stopping criteria. It was shown that the analysis based on RMT helps to improve the convergence and learning speed of neural networks as well as to improve accuracy and reduce the computational complexity of training algorithms in DNNs.

However, there are many open questions in applying RMT to DNNs. For example, RMT often relies on strong simplifying assumptions, such as independence or Gaussianity of network inputs and weights, which may not hold in practice. It is not clear how to extend the RMT analysis to more complex settings, such as non-linear activations, non-i.i.d. data, stochastic gradient descent, and convolutional and recurrent neural networks, etc. There is a need for further research and development of RMT methods that can expand their applicability to more general and realistic DNN scenarios.

In this project we will study spectral properties of random matrix ensembles arising in the analysis of DNNs analytically and numerically, and develop mathematical techniques for improving accuracy and reducing the computational complexity of training algorithms in DNNs. We will address such questions as increasing the efficiency of training by pruning DNN parameters, finding an optimal weight initialization for untrained DNNs, determining training stopping criteria, studying nonlinearity's effect on DNN learning speed, and justification of numerical and approximation methods and error estimates. The expected results will provide new insights into the theoretical foundations of DNNs and will also develop new tools and techniques of RMT that are specifically aimed at applications to DNNs.