PHD PROJECT:
RANDOM MATRICES, HIGH-DIMENSIONAL STATISTICS AND BEYOND

JOHANNES HEINY

My main research interests lie in random matrix theory, high-dimensional statistics and extreme value theory. The dramatic increase and improvement of computing power and data collection devices have triggered the necessity to study and interpret the sometimes overwhelming amounts of data in an efficient and tractable way. Huge data sets arise naturally in wireless communication, finance, natural sciences and genetic engineering. Traditional multivariate analysis relies on the assumption that the dimension p of the data remains fixed and thus is negligible compared to the sample size n. For this reason, results from traditional multivariate analysis are typically not applicable in other regimes. Spurred by these problems, new analysis tools for high-dimensional data were developed in recent years. Often it is assumed that dimension-to-sample-size ratio p/n tends to a positive constant as $n, p \to \infty$.

Project description. The focus of the PhD project is on high-dimensional probability theory and random matrix theory with possible applications in statistics. In particular, the research aims at describing the dependence structure of large data sets which is often studied via sample covariances, correlations, Spearman's ρ or Kendall's τ. Estimating and accurately assessing dependence has become the cornerstone of statistical inference and prediction in high dimension. In this effort, asymptotic probabilistic results from random matrix theory will be employed to obtain statistically meaningful results facilitating effective inference and prediction within the curse of dimensionality domain. Among a variety of important, difficult, and empirically meaningful topics in random matrix theory, the phase transitions and universality phenomena of large random matrices take a particularly prominent place. The underlying methodological flavor of the research aims at combining elegant and sophisticated mathematical theories with a variety of fields of applications.

For example, in modern statistical learning theory, prediction and classification problems are two of the main tasks considered. Different methods, like ridge regression, lasso regression, decision tree, logistic regression, discriminant analysis, etc., have recently been developed in the literature and successfully implemented in practice. Theoretical results, motivating the applicability of the existing approaches, are derived in most cases under the standard asymptotic regime, i.e., when the sample size is considerably larger than the number of predictors in the model. Even though the same procedures are applied in the high-dimensional setting, i.e., when the number of predictors is comparable to the sample size, theoretical justifications are usually not available, and one can face unexpected results.

Within the project, we are going to contribute to statistical learning theory by developing new theoretical findings that explain and justify the applicability of the existing prediction and classification approaches in the high-dimensional setting. This will be achieved by developing new theoretical results in random matrix theory and their applications to the problems at hand.

Date: February 26, 2025.

Aims. The supervisor and the PhD student will choose the PhD topic together to ensure that personal preferences and strengths are accounted for, with the student taking as much initiative as possible. General aims of the PhD project could be to:

- explain the connections between the extremal entries and the spectral properties of large random matrices,
- develop novel statistical tests for high-dimensional data by combining methods from random matrix theory with other fields,
- prove limit theory for functions of the eigenvalues of large random matrices,
- better understand modern machine learning methods for high-dimensional data.

About suitable candidates. This project requires a **good knowledge of probability theory and statistics**, as well as self-motivation, enthusiasm and the willingness to carry out significant research within a lively area of modern mathematics. Some expertise in one or more of the following related subjects can be a minor advantage, but is no way mandatory: high-dimensional statistics, random matrix theory, Gaussian approximations, time series, point processes, extreme value theory.

More information about my research can be found at
https://www.su.se/english/profiles/johe3032-1.640812

There are many potential research topics suitable for PhD studies in this field. Please do not hesitate to contact me at johannes.heiny@math.su.se if you have any questions or want to know about concrete research questions for a PhD project.

Department of Mathematics, Stockholm University, Albano hus 1, 10691 Stockholm, Sweden
Email address: johannes.heiny@math.su.se