Workshop on High-Dimensional Covariance Operators and their Applications

September 12-13, 2019 in Berlin

Abstracts

The Hanson-Wright inequality in Banach Spaces

Radosław Adamczak
University of Warsaw

I will discuss two-sided bounds for moments and tails of Banach space valued quadratic forms in Gaussian random variables. I will state a natural conjecture and show that it holds up to additional logarithmic factors. Moreover in a certain class of Banach spaces (including L_r-spaces) these logarithmic factors may be eliminated. As a corollary one can obtain upper bounds for tails and moments of quadratic forms in subgaussian random variables, which extend the Hanson-Wright inequality (providing such bounds for real-valued quadratic forms). If time permits I will also briefly discuss results for more general Gaussian polynomials with vector valued coefficients. Based on joint work with Rafal Latala and Rafal Meller.

Statistical Optimality of Stochastic Gradient Descent on Hard Learning Problems through Multiple Passes

Francis Bach
INRIA, École normale supérieure Paris

We consider stochastic gradient descent (SGD) for least-squares regression with potentially several passes over the data. While several passes have been widely reported to perform practically better in terms of predictive performance on unseen data, the existing theoretical analysis of SGD suggests that a single pass is statistically optimal. While this is true for low-dimensional easy problems, we show that for hard problems, multiple passes lead to statistically optimal predictions while single pass does not; we also show that in these hard models, the optimal number of passes over the data increases with sample size. In order to define the notion of hardness and show that our predictive performances are optimal, we consider potentially infinite-dimensional models and notions typically associated to kernel methods, namely, the decay of eigenvalues of the covariance matrix of the features and the complexity of the optimal predictor as measured through the covariance matrix. We illustrate our results on synthetic experiments with non-linear kernel methods and on a classical benchmark with a linear model. (Joint work with Loucas Pillaud-Vivien and Alessandro Rudi).

Modern machine learning and kernel machines

Mikhail Belkin
Ohio State University

Random hyperplane tessellations and robust one-bit compressed sensing

Sjoerd Dirksen
Utrecht University

Compressed sensing is a recent paradigm in signal processing that aims to design efficient measurement schemes by exploiting low-complexity structures in signal sets, such as sparsity in a suitable basis. In the traditional compressed sensing literature, it is implicitly assumed that one has direct access to noisy analog linear measurements of an (unknown) signal. In reality, these analog measurements need to be quantized to a finite number of bits before they can be transmitted, stored, and processed. In the emerging theory of quantized compressed sensing it is studied how to jointly design a quantizer, measurement procedure, and reconstruction algorithm in order to accurately recover low-complexity signals.

In my talk I will consider the popular one-bit compressed sensing model, in which each linear analog measurement is quantized to a single bit. I will explain two methods to reconstruct low-complexity signals from their one-bit measurements and present bounds on the number of measurements that is sufficient to guarantee accurate reconstruction in the case that the measurement matrix is either heavy-tailed or a circulant matrix generated by a random vector. The reconstruction results that I will present are optimal in several cases and very robust to noise on the analog measurements as well as to bit corruptions occurring in the quantization process. These results rely on showing that the tessellation generated by a small number of random hyperplanes can be used to approximate Euclidean distances between any two points in a given set.

Based on joint work with Shahar Mendelson (ANU Canberra)

Sample Eigenstructure: A Random Matrix Perspective

Holger Kösters
Universität Rostock

In my talk I will survey some recent developments in random matrix theory concerning the eigenvalues and eigenvectors of high-dimensional sample covariance matrices, with particular emphasis on spiked covariance models.

Linear functionals in PCA and spectral clustering

Matthias Löffler
University of Cambridge

Considering a Gaussian covariance model, I will explain when linear functionals of the first sample eigenvector are biased. This even holds in low-dimensional regimes with p/n going to 0. If there is bias, I will discuss how to correct for the bias to obtain asymptotically unbiased Gaussian estimators with optimal variance. As an example for application of this theory, I will consider spectral clustering in a simple Gaussian mixture model with centres θ and $-\theta$. I will show how the misclustering error can be related to linear functionals of the first eigenvector of the empirical "covariance". This yields an exponential misclustering error rate with minimax optimal constant, even when $p \gg n$. This is based on joint works with V. Koltchinskii, R. Nickl, A.Y. Zhang and H.H. Zhou.

Nonparametric estimation of the ability density in the Mixed-Effect Rasch Model

Alexander Meister
Universität Rostock

The Rasch model is widely used in the field of psychometrics when n persons under test answer m questions and the score, which describes the correctness of the answers, is given by a binary (n, m)-matrix. We consider the Mixed-Effect Rasch Model, in which the persons are chosen randomly from a huge population. The goal is to estimate the ability density of this population under nonparametric constraints, which turns out to be a statistical linear inverse problem with an unknown but estimable operator. Based on our previous result on asymptotic equivalence to a two-layer Gaussian model, we construct an estimation procedure and study its asymptotic optimality properties as n tends to infinity, as does m, but moderately with respect to n. Moreover numerical simulations are provided.

Extreme value theory for the entries of the sample covariance matrix

Thomas Mikosch
University of Copenhagen

We consider the point process of the off-diagonal entries of a sample covariance matrix based on a $p \times n$-dimensional data matrix with iid entries. Under moment conditions, we show that this point process converges to a Poisson process on the real line with mean measure induced by the negative logarithm of the Gumbel distribution. The basis for this result are precise mutivariate large deviation results for independent random walks. This is joint work with Johannes Heiny and Jorge Yslas.

Moment inequalities for matrix-valued U-statistics of order 2

Stanislav Minsker
University of Southern California

We present Rosenthal-type moment inequalities for matrix-valued U-statistics of order 2; the sample covariance matrix is a well-known example of such an object. As a corollary, we obtain new matrix concentration inequalities for U-statistics. One of our main technical tools, a version of the non-commutative Khintchine inequality for the spectral norm of the Rademacher chaos, could be of independent interest. The talk is based on a joint work with Xiaohan Wei.

Robust sparse covariance estimation for elliptical data by thresholding Tyler's M-estimator

Boaz Nadler
Weizmann Institute of Science

Estimating a high-dimensional sparse covariance matrix from a limited number of samples is a fundamental problem in contemporary data analysis. Most proposals to date, however, are not robust to outliers or heavy tails. Towards bridging this gap, we consider estimating a sparse shape matrix from n samples following a possibly heavy tailed elliptical distribution. We propose estimators based on thresholding either Tyler's M-estimator or its regularized variant. We derive bounds on the difference in spectral norm between our estimators and the shape matrix

in the joint limit as the dimension p and sample size n tend to infinity with p/n tending to a constant. These bounds are minimax rate-optimal. Results on simulated data support our theoretical analysis.

Eigenvalues and eigenspaces under random perturbation

Sean O'Rourke
University of Colorado Boulder

When a Hermitian matrix is slightly perturbed, by how much can its eigenvalues and eigenvectors change? Classical (deterministic) theorems, such as those by Davis-Kahan and Weyl, answer this question by giving tight estimates for the worst-case scenario. In this talk, I will consider the case when the perturbation is random. In this setting, better estimates can be achieved when the matrix is approximately low rank. This talk is based on joint work with Van Vu and Ke Wang.

Procrustes Metrics and Optimal Transport of Covariance Operators

Victor Panaretos
École polytechnique fédérale de Lausanne

Covariance operators are fundamental in functional data analysis, providing the canonical means to analyse functional variation via the celebrated Karhunen-Loève expansion. These operators may themselves be subject to variation, for instance in contexts where multiple functional populations are to be compared. Statistical techniques to analyse such variation are intimately linked with the choice of metric on covariance operators, and the intrinsic infinite-dimensionality of these operators. We will describe the manifold-like geometry of the space of trace-class infinite-dimensional covariance operators and associated key statistical properties, under the recently proposed infinite-dimensional version of the Procrustes metric. In particular, we will identify this space with that of centred Gaussian processes equipped with the Wasserstein metric of optimal transportation. The identification allows us to provide a description of those aspects of the geometry that are important in terms of statistical inference, and establish key properties of the Fréchet mean of a random sample of covariances, as well as generative models that are canonical for such metrics. The latter will allow us to draw connections with the problem of registration of warped functional data. Based on joint work with V. Masarotto (Leiden) and Y. Zemel (Göttingen).

Bootstrapping linear spectral statistics of high-dimensional covariance matrices

Angelika Rohde
Universität Freiburg

We introduce a new ,$(m, mp/n)$ out of (n, p)'-sampling with replacement bootstrap for linear spectral statistics of high-dimensional sample covariance matrices based on n independent p-dimensional random vectors. For a large class of population covariance matrices satisfying the so-called representative subpopulation condition, this fully nonparametric bootstrap is shown to be consistent in the high-dimensional scenario $p/n \to c \in (0, \infty)$ iff $m^2/n \to 0$. This is in sharp contrast to the inconsistency of the classical sampling with replacement bootstrap in the high-dimensional scenario. The talk is based on a joint work with Holger Dette (Ruhr-Universität Bochum).