Workshop:

Structural inference in high-dimensional models

The conference will be held at the National Research University Higher School of Economics (Myasnitskaya, 20) from September 5 to September 8, 2018 in the room 101.

Program (v. 3.09.2018)

September 5

	Speaker	Title
13:00-13:50	Registration	
13:50-14:00	Welcome	
14:00-15:00	Francis Bach	Large-scale machine learning and convex optimization 1
15:00-15:30	**Coffee break**	
15:30-16:10	Alain Durmus	Bridging the Gap between Constant Step Size Stochastic Gradient Descent and Markov Chains
16:15-16:55	Blazej Miasojedow	Analysis of Langevin Monte Carlo via convex optimization
16:55-17:15	**Tea break + Photo**	
17:15-17:55	Denis Belomestny	Variance reduction for MCMC via regression based control variates
18:00-18:40	Natalia Smorodina	Approximation of an evolution operator by mathematical expectations of functionals of Poisson random fields.
18:45-21:00	**Welcome party + Posters, Myasnitskaya, 9/11, room 325**	

September 6

	Speaker	Title
10:00-11:00	Arnak Dalalyan	On robust estimation of a high dimensional parameter 1
11:00-11:30	**Coffee break**	
11:30-12:10	Yury Golubev	Bayes tests in sparse vectors detection
12:15-12:55	Mathias Trabs	Diffusion matrix estimation for high-dimensional Lévy processes
12:55-14:15	**Lunch**	
14:15-15:15	Francis Bach	Large-scale machine learning and convex optimization 2
15:25-16:05	Gerard Biau	Some theoretical properties of GANs
16:05-16:35	**Tea break**	
16:35-17:15	Liza Rebrova	Constructive regularization of random matrix norm
17:20-18:00	Alexander Tikhomirov	On the distribution of maximum of large number of linear and quadratic forms in independent random variables
18:00-19:00	**Discussions**	

September 7

	Speaker	Title
10:00-11:00	Arnak Dalalyan	On robust estimation of a high dimensional parameter 2
11:00-11:30	**Coffee break**	
11:30-12:10	Marko Cuturi	Regularized optimal transport
12:15-12:55	Jean-Michel Loubes	Fairness in Machine Learning Algorithms using Optimal Transport
12:55-14:15	**Lunch**	
14:15-14:55	Clement Levrard	Convergence rates and topological inference for manifold estimation
15:00-15:40	Quentin Paris	Rates of convergence of empirical barycenters in metric spaces
15:40-16:10	**Tea break**	
16:10-16:50	Andrey Zaitsev	Rare events and Poisson point processes
16:55-17:35	Dmitry Zaporozhets	Angles of Random Polytopes

September 8

	Speaker	Title
10:00-10:40	Steve Oudot	The pre-image problem from a topological perspective
10:45-11:25	Evgeny Burnaev	Quadrature-based features for kernel approximation
11:25-11:50	**Coffee break**	
11:50-12:30	Anna Ben-Hamou	Estimating graph parameters with random walks
12:35-13:15	Maxim Panov	Parameter Estimation in Mixed Membership Stochastic Block Models
13:20-	**Lunch**	

Abstracts

Francis Bach (INRIA - Ecole Normale Supérieure)
e-mail: francis.bach@inria.fr

Large-scale machine learning and convex optimization

Many statistics, machine learning and signal processing problems are traditionally cast as convex optimization problems. A common difficulty in solving these problems is the size of the data, where there are many observations ("large n") and each of these is large ("large p"). In this setting, online algorithms such as stochastic gradient descent which pass over the data only once, are usually preferred over batch algorithms, which require multiple passes over the data. Given n observations/iterations, the optimal convergence rates of these algorithms are $O(1/\sqrt{n})$ for general convex functions and reaches $O(1/n)$ for strongly-convex functions.

Arnak Dalalyan (CREST-ENSAE)
e-mail: arnak.dalalyan@ensae.fr

On robust estimation of a high dimensional parameter

This tutorial will present some recent advances on the problem of estimating a high dimensional mean vector or a covariance matrix in the case where the sample contains outliers. The focus will be on finite sample results providing rate optimal risk bounds. We will review several approaches including Huber's estimator, coordinatewise and geometric medians, Tukey's median, iterative group-soft-thresholding, etc. A particular emphasis will be put on the analysis of estimators that have polynomial computational complexity.

Main references:

1) Olivier Collier, Arnak S. Dalalyan (2017). Rate-optimal estimation of p-dimensional linear functionals in a sparse Gaussian model.
2) Guillaume Lecue, Matthieu Lerasle (2017). Learning from MOM's principles: Le Cam's approach.
3) M. Chen, C. Gao, Z. Ren (2015) Robust Covariance and Scatter Matrix Estimation under Huber's Contamination Model.
4) Samuel Balmand, Arnak S. Dalalyan (2015). Convex programming approach to robust estimation of a multivariate gaussian model.
5) M. Lerasle, R. I. Oliveira (2011). Robust empirical mean Estimators.

Denis Belomestny (HSE; University of Duisburg-Essen)
e-mail: denis.belomestny@uni-due.de

Variance reduction for MCMC via regression based control variables

In this talk we focus on the so-called unadjusted Langevin algorithm (UL) which got much attention recently. We propose a generic variance reduction method for UL algorithms, which is purely empirical and doesn't require knowledge of the norming constants. Moreover, we don't need to assume sampling under the underlying invariant distribution. We rigorously analyse the convergence of the method and study its complexity. It is shown that variance reduced UL method outperforms the standard UL algorithm in terms of complexity.

Anna Ben-Hamou (Sorbonne University, Paris)
e-mail: anna.ben-hamou@upmc.fr

Estimating graph parameters with random walks

Let G be a fixed but unknown graph, which can only be explored locally, through random walks started at a given vertex. With such a procedure, is it possible to infer some graph parameters, such as its size, and how fast? In this talk, we will present some simple algorithms based on the number of intersections between random walk' trajectories. These algorithms improve on previous methods by various authors, and we will show that their time complexity is optimal.

Gérard Biau (Sorbonne University)
e-mail: gerard.biau@upmc.fr

Some theoretical properties of GANs

Generative Adversarial Networks (GANs) are a class of generative algorithms that have been shown to produce state-of-the art samples, especially in the domain of image creation. The fundamental principle of GANs is to approximate the unknown distribution of a given data set by optimizing an objective function through an adversarial game between a family of generators and a family of discriminators. We offer a better theoretical understanding of GANs by analyzing some of their mathematical and statistical properties. We study the deep connection between the adversarial principle underlying GANs and the Jensen-Shannon divergence, together with some optimality characteristics of the problem. An analysis of the role of the discriminator family via approximation arguments is also provided. In addition, taking a statistical point of view, we study the large sample properties of the estimated distribution and prove in particular a central limit theorem. Some of our results are illustrated with simulated examples. Joint work with B. Cadre, M. Sangnier, and U. Tanielian.

Evgeny Burnaev (Skoltech; HSE)
e-mail: e.burnaev@skoltech.ru

Quadrature-based features for kernel approximation

We consider the problem of improving kernel approximation via randomized feature maps. These maps arise as Monte Carlo approximation to integral representations of kernel functions and scale up kernel methods for larger datasets. We propose to use more efficient numerical integration technique to obtain better estimates of the integrals compared to the state-of-the-art methods. Our approach allows the use of information about the integrand to enhance approximation and facilitates fast computations. We derive the convergence behavior and conduct an extensive empirical study that supports our hypothesis.

Marko Cuturi (CREST, ENSAE)
e-mail: marco.cuturi@ensae.fr

Regularized optimal transport

TBA

Alain Durmus (ENS Paris-Saclay)
e-mail: alain.durmus@cmla.ens-cachan.fr

Bridging the Gap between Constant Step Size Stochastic Gradient Descent and Markov Chains

We consider the minimization of an objective function given access to unbiased estimates of its gradient through stochastic gradient descent (SGD) with constant step-size. While the detailed analysis was only performed for quadratic functions, we provide an explicit asymptotic expansion of the moments of the averaged SGD iterates that outlines the dependence on initial conditions, the effect of noise and the step-size, as well as the lack of convergence in the general (non-quadratic) case. For this analysis, we bring tools from Markov chain theory into the analysis of stochastic gradient and create new ones (similar but different from stochastic MCMC methods). We then show that Richardson-Romberg extrapolation may be used to get closer to the global optimum and we show empirical improvements of the new extrapolation scheme. (joint work with Francis Bach and Aymeric Dieuleveut).

Yuri Golubev (CNRS)
e-mail: gueorgui.goloubev@amu-univ.fr

Bayes tests in sparse vectors detection

In this talk we compute and analyze limit distributions of the Bayes test statistics related to sparse vectors detection. More precisely, the goal is to test simple hypothesis $H_0 : \|S\| = 0$ versus compound alternative $H_1 : \|S\| > 0$ based on the observations

$$Y_i = S_i \cdot \mathbf{1}(i \in \tau) + \xi_i, \quad i = 1, 2, \dots,$$

where $S = (S_1, S_2, \dots)^\top \in \mathbb{R}^\infty$ is an unknown vector, ξ_i are i.i.d. $\mathcal{N}(0, \sigma^2)$, $\tau = \{\tau_1, \dots, \tau_p\}$ is an unknown random multi-index with i.i.d. components and a known a priori distribution $\mathbf{P}\big\{\tau_k = i\big\} = \bar{\pi}_i,\ i \in \mathbb{Z}^+$.

Assuming that entropy $H(\bar{\pi})$ of $\bar{\pi}$ tends to infinity, we compute limit distributions of the Bayes test statistics. This permits to compare Bayes tests with the maximum a posteriori probability test and thus to find out when and in what sense the Bayes test is optimal.

It is shown also that the limit distribution of the Bayes test for $p = 1$ is characterized by the following property: let ζ_i be i.i.d. random variables which follow the Bayes limit law, then

$$\sum_{i=1}^{\infty} \lambda_i \zeta_i \stackrel{\mathbf{P}}{=} H(\lambda) + \zeta_1 \text{ for any } \lambda_i \geq 0, \text{ such that } \sum_{i=1}^{\infty} \lambda_i = 1.$$

Clement Levrard (University Paris-Diderot)
e-mail: clement.levrard@lpsm.paris

Convergence rates and topological inference for manifold estimation

In high-dimensional settings, a non-linear extension of the sparsity assumption can be to assume that data points lie on a low-dimensional structure that can be modeled as a submanifold of the ambient space. Such an assumption can provide improved convergence rates for statistical procedures such as kernel regression whenever features are assumed to lie on a low-dimensional submanifold. This talk will focus on the estimation of the underlying submanifold itself. In the first part I will expose what precision can be expected in terms of Hausdorff distance, discussing the influence of sample size, dimensions (ambient and intrinsic) as well as regularity of the submanifold. An interesting conclusion is that the ambient dimension plays actually no role. In the second part of the talk I will present a feasible submanifold estimator, based on the Tangential Delaunay Complex, that achieves both optimal convergence rate over a low-regularity class of submanifolds and exact topological recovery with high probability. This is joint work with Eddie Aamari (CNRS).

Jean-Michel Loubes (University of Toulouse)
e-mail: loubes@math.univ-toulouse.fr

Fairness in Machine Learning Algorithms using Optimal Transport

Statistical algorithms are usually helping in making decisions in many aspects of our lives. But, how do we know if these algorithms are biased and commit unfair discrimination of a particular group of people, typically a minority? Fairness is generally studied in a probabilistic framework where it is assumed that there exists a protected variable, whose use as an input of the algorithm may imply discrimination. There are different definitions of Fairness in the literature. In this paper we focus on two of them which are called Disparate Impact (DI) and Balanced Error Rate (BER). Both are based on the outcome of the algorithm across the different groups determined by the protected variable. The relationship between these two notions is also studied. The goals of this paper are to detect when a binary classification rule lacks fairness and to try to fight against the potential discrimination attributable to it. This can be done by modifying either the classifiers or the data itself. Our work falls into the second category and modifies the input data using optimal transport theory. We provide a new way of achieving fairness using OT and Wasserstein Barycenters.

Blazej Miasojedow (Institute of Mathematics, Polish Academy of Sciences)
e-mail: bmia@mimuw.edu.pl

Analysis of Langevin Monte-Carlo via convex optimization

We provide new insights on the Unadjusted Langevin Algorithm. We show that this method can be formulated as a first order optimization algorithm of an objective functional defined on the Wasserstein space of order 2. Using this interpretation and techniques borrowed from convex optimization, we give a non-asymptotic analysis of this method to sample from log concave smooth target distribution. Our proofs are then easily extended to the Stochastic Gradient Langevin Dynamics, which is a popular extension of the Unadjusted Langevin Algorithm. Finally, this interpretation leads to a new methodology to sample from a non-smooth target distribution, for which a similar study is done.

Steve Oudot (Inria; Ecole Polytechnique)
e-mail: steve.oudot@inria.fr

The pre-image problem from a topological perspective

This talk will be a review of the efforts of the Topological Data Analysis (TDA) community to tackle the preimage problem. The main focus will be on recent attempts to derive left and right inverses for the TDA operator. While this line of work is still in its infancy, the hope on the long run is to use such inverses for feature interpretation. The mathematical tools involved in the analysis come mainly from metric geometry and sheaf theory, and specific pointers will be given in the course of the exposition.

Maxim Panov (Skoltech; HSE)
e-mail: panov.maxim@gmail.com

Parameter Estimation in Mixed Membership Stochastic Block Models

Community detection is an important problem in modern network analysis. It has wide applications in analysis of social and biological networks, designing network protocols and many other areas. Recently, much attention has been paid to detection of overlapping communities, where each node in a network may belong to multiple communities. We note that the majority of overlapping community detection methods come with no guarantees on their performance. This paper considers the parameter estimation problem in Mixed Membership Stochastic Block Model (MMSB), which is a quite general instance of random graph model allowing for overlapping community structure. We consider algorithm successive projection overlapping clustering (SPOC) which combines the ideas of spectral clustering and geometric approach for separable non-negative matrix factorization in MMSB. The algorithm starts from the spectral embedding based on the adjacency matrix of the graph, then finds nearly pure nodes via successive projection algorithm (SPA) and finally reconstructs community memberships via least-squares fit. SPOC is provably consistent under MMSB with general conditions on the parameters of the model and it is also shown to perform well experimentally in comparison to other algorithms.

In this work, we provide lower bounds on the rates of parameter estimation in MMSB and discuss the question of minimax optimiality in the considered problem.

Quentin Paris (HSE)
e-mail: qparis@hse.ru

Convergence rates for empirical barycenters in metric spaces

The talk presents rates of convergence for empirical barycenters of a Borel probability measure on a metric space under general conditions. Our main assumption connects ideas from metric geometry to the theory of empirical processes. This assumption is discussed in two meaningful scenarios. The first one is a geometrical constraint on the underlying space referred to as (k, α)-convexity, compatible with a positive upper curvature bound in the sense of Alexandrov. The second scenario considers the case of a nonnegatively curved space on which geodesics, emanating from a barycenter, can be extended. Our results are discussed in the context of Wasserstein spaces.

Liza Rebrova (UCLA)
e-mail: rebrova@math.ucla.edu

Constructive regularization of random matrix norm

In our work with R. Vershynin we have shown that if (and only if) a n*n random matrix A have i.i.d. entries with zero mean and finite second moment, the operator norm of A can be regularized to the optimal order $O(\sqrt{n})$ locally (i. e., by zeroing out a small submatrix of A). A natural question arising from this result is how to find the submatrix to be deleted. The main goal of this talk is to discuss a recent local norm regularization algorithm that works with high probability and reduces the matrix norm to the almost optimal order $O(\sqrt{n \log\log n})$. The new approach to the norm regularization extends techniques developed for the regularization of graph adjacency (Bernoulli) matrices in the works of Feige, Ofek, and Le, Levina, Vershynin.

Natalia Smorodina (PDMI RAS)
e-mail: smorodina@pdmi.ras.ru

Approximation of an evolution operator by mathematical expectations of functionals of Poisson random fields

Consider an operator $H = -\frac{1}{2}\frac{d^2}{dx^2} + V(x)$ on the domain $W_2^2(\mathbb{R})$. Suppose that the potential V is real-valued and bounded that implies that the operator H is self-adjoint. The operator family e^{-itH} is a group of unitary operators in $L_2(\mathbb{R})$. The operator e^{-itH} maps a function $\varphi \in W_2^2(\mathbb{R})$ into the Cauchy problem solution $u(t, x)$ for the Schrödinger equation $i\frac{\partial u}{\partial t} = Hu$ with an initial function $u(0, x) = \varphi(x)$ (more details see in[1]). It is well known that a solution of the Cauchy problem for the heat equation $\frac{\partial u}{\partial t} = -Hu,\ u(0, x) = \varphi(x)$ admits a probabilistic representation in the form of an expectation of a Wiener process functional (Feynmann- Kac formula)

$$u(t, x) = e^{-tH}\varphi(x) = \mathbf{E}\left[\varphi(x + w(t))e^{-\int_0^t V(x+w(\tau))\,d\tau}\right], \tag{1}$$

where $w(t)$ is a standard Wiener process. Formula (1) means that that one can simulate the evolution of initial function φ under the heat semigroup e^{-tH} generating the Wiener process trajectories.

In our talk a similar approach is developed for the operator e^{-itH}. Namely, we construct a family Q_ε^t of operators in $L_2(\mathbb{R})$, depending on an additional parameter $\varepsilon > 0$ and possessing the following properties

1) for every $\varepsilon > 0$ the family Q_ε^t is a semigroup, i.e. $Q_\varepsilon^{t+s} = Q_\varepsilon^t Q_\varepsilon^s$,

2) the operator norm of the operator Q_ε^t is not greater than 1,

3) Q_ε^t is defined as expectation of a Poisson point field functional,

4) as $\varepsilon \to 0$ operators Q_ε^t approximate the operator e^{-itH} in strong operator convergence sense that is for every $\varphi \in L_2(\mathbb{R})$ we have $\|Q_\varepsilon^t \varphi - e^{-itH}\varphi\|_2 \longrightarrow 0$.

The above properties yield that that one can simulate the evolution of initial function φ under the group e^{-itH} generating the Poisson point field trajectories. It is worth to mention that. the square of wave function modulus is a density of a probability distribution. A wave function evolution generates an evolution of probability distribution density which is usually called a "quantum random walk". The suggested approach gives a theoretic possibility to simulate the "quantum random walk" by classical statistical technique. A particular case of the above construction can be found in [2].

REFERENCES

(1) *Glimm J., Jaffe A.* Quantum Physics. A Functional Integral Point of View, Springer-Verlag, New York Heidelberg Berlin, 1987.

(2) *Ibragimov I.A., Smorodina N.V., Faddeev M.M.* On a limit theorem related to probabilistic representation of the Cauchy problem solution for Schrödinger equation. Zap. nauchn. semin. PDMI, 2016, v.454, pp. 158-176, (in russian).

Alexander Tikhomirov (Komi SC RAS)

e-mail: antikhom51@gmail.com

On the distribution of maximum of large number of linear and quadratic forms in independent random variables

TBA

Mathias Trabs (Universitat Hamburg)

e-mail: mathias.trabs@uni-hamburg.de

Diffusion matrix estimation for high-dimensional Lévy processes

The estimation of the diffusion matrix Σ of a possibly time-changed Lévy process is studied, based on discrete observations of the process with a fixed distance. We investigate a high-dimensional regime where the dimension d may grow with sample size n. Assuming first that the diffusion matrix sparse, we construct thresholding estimators based on a spectral approach. Minimax optimal convergence rates are proven which are logarithmic in $n/(\log d)$. In a second scenario, we impose a low-rank condition on Σ and construct a weighted least-squares estimator in spectral domain with nuclear norm penalisation. Also in this case we prove oracle inequalities and derive convergence rates for the diffusion matrix estimator. The talk is based on joint works with Denis Belomestny and Alexandre B. Tsybakov.

Andrei Zaitsev (PDMI RAS)

e-mail: zaitsev@pdmi.ras.ru

Rare events and Poisson point processes

The aim of the present talk is to show that the results obtained earlier on the approximation of distributions of sums of independent terms by the accompanying compound Poisson laws may be interpreted as rather sharp quantitative estimates for the closeness between the sample containing independent observations of rare events and the Poisson point process which is obtained after a Poissonization of the initial sample.

Dmitry Zaporozhets (PDMI RAS)

e-mail: zap1979@gmail.com

Angles of Random Polytopes

We will consider some problems on calculating the average angles of random polytopes. Some of them are open.

Posters

- Leonid Iosipoi, Comparing Least Squares and Empirical Variance Minimization in Variance Reduction
- Alexey Kroshnin, Central Limit Theorem for Wassserstein Barycenters of Gaussian Measures
- Anna Husakova, TBA
- Nikita Puchkin, Pointwise adaptation via stagewise aggregation of local estimates for multiclass classification
- Mikhail Monakhov, Second Order Chebyshev-Edgeworth and Cornish-Fisher Expansions for Distributions of Statistics Constructed from Samples with Random Sizes
- Maria Danshina, Limit distribution of vertex degrees in generalized random graph
- Yury Tavyrikov, TBA
- Valentina Shumovskaya, Towards Hypothesis Testing for Random Graphs with Community Structure
- Gabriela Ciolek, High dimensional central limit theorem for one-dependent data