TITLES & ABSTRACTS

Chiara Amorino (UPF Barcelona)

Title: Polynomial rates via deconvolution for nonparametric estimation in McKean-Vlasov SDEs

Abstract: This paper investigates the estimation of the interaction function for a class of McKean-Vlasov stochastic differential equations. The estimation is based on observations of the associated particle system at time T, considering the scenario where both the time horizon T and the number of particles N tend to infinity. Our proposed method recovers polynomial rates of convergence for the resulting estimator. This is achieved under the assumption of exponentially decaying tails for the interaction function. Additionally, we conduct a thorough analysis of the transform of the associated invariant density as a complex function, providing essential insights for our main results.

Juntong Chen (University of Twente)

Title: A statistical analysis of an image classification problem

Abstract: The availability of massive image databases resulted in the development of scalable machine learning methods such as convolutional neural network (CNNs) filtering and processing these data. While the very recent theoretical work on CNNs focuses on standard nonparametric denoising problems, the variability in image classification datasets does, however, not originate from additive noise but from variation of the shape and other characteristics of the same object across different images. To address this problem, we consider a supervised classification problem for object detection on grayscale images. While from the function estimation point of view, every pixel is a variable and large images lead to high-dimensional function recovery tasks suffering from the curse of dimensionality, increasing the number of pixels in our image deformation model enhances the image resolution and makes the object classification problem easier. We propose and theoretically analyze two different procedures. The first method estimates the image deformation by using inverse mappings. Under a minimal separation condition, it is shown that perfect classification is possible. The second method fits a CNN to the data. We establish the theoretical performance for the misclassification error, which depends on the sample size and the number of pixels.

Reda Chhaibi (Université de Toulouse)

Title: "How to estimate a covariance matrix? (In large dimensions)"

Abstract: Consider a sequence of n observations of iid vectors in d-dimensional space. The classical covariance estimator is valid when d is fixed and n is large. However, random matrix theory and free probability suggest that this estimator is no longer valid when dand n are of comparable size. This is just a reinterpretation of the classical Marchenko-Pastur theorem, which we shall explain. A non-trivial "bias" emerges in high-dimensional settings. This phenomenon poses significant challenges, particularly as covariance matrix estimation is a fundamental component in many statistical methods, such as Principal Component Analysis (PCA).

In this presentation, we will explore effective estimation in high dimensions.
Specifically, we will discuss:

- Statistical aspects, including a Cramer-Rao bound for a motivated model.
- Computational aspects, based on complex analysis, orthogonal polynomials and likelihood maximization.

This is work in collaboration with F. Gamboa, S. Kammoun, and M. Velasco.

Christa Cuchiero (University of Vienna)

Title: Conditional polynomial McKean-Vlasov SDEs

Abstract: We study a new class of McKean-Vlasov stochastic differential equations (SDEs), possibly with common noise, applying the theory of time-inhomogeneous polynomial processes. The drift and volatility coefficients of these SDEs depend on the state variables themselves as well as their conditional moments in a way that mimics the standard polynomial structure. Such conditional McKean-Vlasov SDEs arise for instance in the context of stochastic portfolio theory when interacting particle systems are used to model the capital distribution curves. Our approach leads to new results on the existence and uniqueness of solutions to such conditional McKean-Vlasov SDEs which were, to the best of our knowledge, not obtainable using standard methods. Indeed, the special form of the characteristics causes the conditional moments to become an autonomous standard Ito-SDE driven by the common noise. Provided that a pathwise unique solution exists to this SDE, we can establish an existence, uniqueness and tractability theory that covers a large class of conditional McKean Vlasov SDEs beyond the standard conditions of Lipschitz continuity and uniform ellipticity. In the case without common noise the SDE reduces to a non-linear ODE, which determines the moments of the McKean-Vlasov SDE. As a by-product, this also yields new results on the existence and uniqueness of global solutions to certain ODEs. The approach can also be extended to an affine structure where the drift and volatility may depend more generally on the characteristic function of the process' marginals and thus on the full law and not only on the moments.

The talk is based on joint work with Janka Möller.

Chao Gao (University of Chicago)

Title: Computational lower bounds for graphon estimation via low-degree polynomials

Abstract: Graphon estimation has been one of the most fundamental problems in network analysis and has received considerable attention in the past decade. From the statistical perspective, the minimax error rate of graphon estimation has been established by Gao et al (2015) for both stochastic block model (SBM) and nonparametric graphon estimation. The statistical optimal estimators are based on constrained least squares and have computational complexity exponential in the dimension. From the computational perspective, the best-known polynomial-time estimator is based on universal singular value thresholding (USVT), but it can only achieve a much slower estimation error rate than the minimax one. It is natural to wonder if such a gap is essential. The computational optimality of the USVT or the existence of a computational barrier in graphon estimation has been a long-standing open problem. In this work, we take the first step towards it and provide rigorous evidence for the computational barrier in graphon estimation via low-degree polynomials. Specifically, in both SBM and nonparametric graphon estimation, we show that for low-degree polynomial estimators, their estimation error rates cannot be significantly better than that of the USVT under a wide range of parameter regimes. Our results are proved based on the recent development of low-degree polynomials by Schramm and Wein (2022), while we overcome a few key challenges in applying it to the general graphon estimation problem. By leveraging our main results, we also provide a computational lower bound on the clustering error for community detection in SBM with a growing number of communities and this yields a new piece of evidence for the conjectured Kesten-Stigum threshold for efficient community recovery.

Elisabeth Gassiat (Orsay Paris)

Title: Fundamental limits for late change-point detection under the preferential attachment random graph model.

Abstract: We consider the problem of late change-point detection under the preferential attachment random graph model with time dependent attachment function. This can be formulated as a hypothesis testing problem where the null hypothesis corresponds to a preferential attachment model with a constant affine attachment parameter δ_0 and the alternative corresponds to a preferential attachment model where the affine attachment parameter changes from δ_0 to δ_1 at a moment $\tau_n = n - \lfloor c n^{\gamma}\rfloor$ where $c > 0$, $\gamma\in (0, 1)$ and n is the size of the graph. It was conjectured that when observing only the unlabeled graph, detection of the change is not possible for $\gamma\in (0, 1/2)$. In this work, we prove this conjecture for $\gamma\in (0, 1/3)$. We also study change-point detection in the case where the labelled graph is observed and show that change-point detection is possible if and only if $n - \tau_n \xrightarrow[n\rightarrow +\infty]{}+\infty$, thereby exhibiting a strong difference between the two settings.

Sophie Langer (University of Twente)

Title: Understanding dropout in the linear model

Abstract: Overparameterized neural networks have gained significant attention in recent years due to their remarkable ability to achieve high accuracy on complex tasks. However, these networks are prone to overfitting, where they memorize the training data rather than learning the underlying patterns. To address this issue, researchers have developed various regularization schemes. In addition to explicit regularization techniques such as l2- or l1-penalization, algorithmic regularization approaches have been employed. Among them, dropout has emerged as a technique that randomly drops neurons during training, and it has demonstrated its effectiveness in various applications. However, despite its empirical success, a comprehensive theoretical understanding of how dropout achieves regularization is still somewhat limited.

In the case of a linear model, it was shown that under an averaged form of dropout the least squares minimizer performs a weighted variant of l2-penalization. In turn, the heuristic "dropout performs l2-penalization" has even made it in popular textbooks. We challenge this relation by investigating the statistical behavior of iterates generated by gradient descent with dropout. In particular, non-asymptotic convergence rates for the expectation and covariance matrices of the iterates are derived. While in expectation the connection between dropout and l2-penalization can be verified, we show sub-optimality of the asymptotic variance compared to the estimator resulting from direct minimization of averaged dropout. As an illustrative example, we also discuss a simplified variant of dropout, which features much simpler interactions. This talk is based on joint work with Johannes Schmidt-Hieber and Gabriel Clara.

Nicolas Lengert (University of Luxembourg)

Title: Limit theorems for ambit fields observed along curves.

Abstract : This article delves into the asymptotic behavior of power variations of continuous two-dimensional ambit fields observed along a curve in $\R^2$. Specifically, the ambit field under consideration is an integral involving a weight kernel $g:\mathbb{R}^2 \to \mathbb{R}$ and a stochastic intermittency process σ, driven by Gaussian white noise. Our investigation demonstrates that the limit theory for the power variation statistics critically hinges on the behavior of the weight kernel g around 0. We explore two distinct cases: when $g({\bf x})\sim c \|{\bf x}\|^{\alpha}$ and $g({\bf x})\sim c|x_1|^{\alpha} |x_2|^{\alpha}$ as ${\bf x}\to {\bf 0}\in \R^2$. These cases yield markedly different asymptotic theories for power variations. In both instances, we establish the corresponding laws of large numbers and stable central limit theorems.

Karim Lounici (Ecole Polytechnique)

Title: Learning the Infinitesimal Generator of Stochastic Diffusion Processes

Abstract: We address data-driven learning of the infinitesimal generator of stochastic diffusion processes, essential for understanding numerical simulations of natural and physical systems. The unbounded nature of the generator poses significant challenges, rendering conventional analysis techniques for Hilbert-Schmidt operators ineffective. To overcome this, we introduce a novel framework based on the energy functional for these stochastic processes. Our approach integrates physical priors through an energy-based risk metric in both full and partial knowledge settings. We evaluate the statistical performance of a reduced-rank estimator in reproducing kernel Hilbert spaces (RKHS) in the partial knowledge setting. Notably, our approach provides learning bounds independent of the state space dimension and ensures non-spurious spectral estimation. Additionally, we elucidate how the distortion between the intrinsic energy-induced metric of the stochastic diffusion and the RKHS metric used for generator estimation impacts the spectral learning bounds.

Eric Luçon (Université Paris Cité)

Title: The validity of mean-field approximation for dynamics on a random graph

Abstract: The aim of the talk it to consider a system of N diffusions interacting on a (possibly random) graph. An easy instance corresponds to the case of full-connectivity, that is when the graph of interaction is complete (mean-field case). In this case, the empirical measure of the system converges as $N\to\infty$ to the solution of a nonlinear Fokker-Planck equation. The question is then: what happens if one considers an nontrivial graph of interaction, that is no longer complete, possibly (very) diluted ? the coupling is no longer a functional of the empirical measure, but rather of local empirical measures around each vertex. What is the proper condition on the graph of interaction under which the behavior of the system remains as in the mean-field case? We address this question of universality both at the level of the law of large numbers and fluctuations, for a large class of possibly random graphs, including the Erdös-Rényi class. We will show in particular that the dependence of the initial condition w.r.t. the graph is crucial. This is based on joint works with S. Delattre, G. Giacomin, F. Coppini and C. Poquet.

Pierre Monmarche (Sorbonne Paris)

Title: Local convergence rates for mean-field diffusion processes

Abstract: For the granular media equation, or other Wasserstein gradient flows associated to some free energy, the exponential convergence towards a unique global minimizer is known to follow from a suitable non-linear log-Sobolev inequality. However this inequality cannot hold when the free energy admits non-global local minimizers, as in the granular media case in a double-well potential with attractive interaction below the critical temperature. We will discuss how local inequalities can still be established in this context to obtain local convergence rates for initial conditions in a Wasserstein ball centered at local minimizers. From a practical point of view, this implies that the free energy of the interacting particles system associated to the mean-field flow decays quickly below the level of the local minimum, up to some error term. The same analysis works in the kinetic case (i.e. the Vlasov-Fokker-Planck equation). Joint work with Julien Reygner.

Jean Christophe Mourrat (ENS Lyon)

Title: Spin glasses with multiple types

Abstract: Spin glasses are models of statistical mechanics in which a large number of elementary units interact with each other in a disordered manner. In the simplest case, there are direct interactions between any two units in the system, and I will start by reviewing some of the key mathematical results in this context. For modelling purposes, it is also desirable to consider models with more structure, such as when the units are split into two groups, and the interactions only go from one group to the other one. I will then discuss some of the technical challenges that arise in this case, as well as recent progress.

Ivan Nourdin (University of Luxembourg)

Title: Quantitative CLT for deep neural networks

Abstract: I will discuss the asymptotic behavior at initialization of fully connected deep neural networks with Gaussian weights and biases when the widths of the hidden layers go to infinity. The focus of the talk will be on the one-dimensional case and optimal bounds for the total variation distance obtained by means of Stein's method. This is based on a joint work with S. Favaro, B. Hanin, D. Marinucci and G. Peccati.

Gabriel Peyré (ENS Paris)

Title: Transformers are *Universal* In-context Learners

Abstract: Transformers deep networks define "in-context mappings'", which enable them to predict new tokens based on a given set of tokens (such as a prompt in NLP applications or a set of patches for vision transformers). This work studies the ability of these architectures to handle an arbitrarily large number of context tokens. To mathematically and uniformly address the expressivity of these architectures, we consider that the mapping is conditioned on a context represented by a probability distribution of tokens (discrete for a finite number of tokens). The related notion of smoothness corresponds to continuity in terms of the Wasserstein distance between these contexts. We demonstrate that deep transformers are universal and can approximate continuous in-context mappings to arbitrary precision, uniformly over compact token domains. A key aspect of our results, compared to existing findings, is that for a fixed precision, a single transformer can operate on an arbitrary (even infinite) number of tokens. Additionally, it operates with a fixed embedding dimension of tokens (this dimension does not increase with precision) and a fixed number of heads (proportional to the dimension). The use of MLP layers between multi-head attention layers is also explicitly controlled. This is a joint work with Takashi Furuya (Shimane Univ.) and Maarten de Hoop (Rice Univ.).

Patricia Reynaud-Bouret (Université Côte d'Azur)

Title: Theoretical and practical implications of the Kalikow decomposition in the study of neuronal networks: simulation and learning

Abstract: Kalikow decomposition is a decomposition of stochastic processes (usually finite state discrete time processes but also more recently point processes) that consists in picking at random a finite neighborhood in the past and then make a transition in a Markov manner. This kind of approach has been used for many years to prove existence of some processes, especially their stationary distribution. In particular, it allows to prove the existence of processes that model infinite neuronal networks, such as Hawkes like processes or Galvès-Löcherbach processes. But beyond mere existence, this decomposition is a wonderful tool to simulate such network, as an open physical system, that from a computational point of view could be competitive with the most performant brain simulations. Finally this decomposition is also a source of inspiration to understand how local rules at each neuron can make the whole network learn.

Radomyra Shevchenko (Université Côte d'Azur)

Title: Some estimation problems for SDEs with a fractional noise

Abstract: We consider estimators of a deterministic periodic drift function in continuously observed Ornstein-Uhlenbeck type SDEs driven by an additive fractional Brownian noise or its non-Gaussian generalization. We comment on the optimality question for the parametric case, and then construct and compare some projection-based nonparametric estimators. We also address related estimation problems for diagonalisable SPDEs that can be dealt with by means of the spectral approach.

Mathias Trabs (KIT Karlsruhe)

Title: Statistical guarantees for stochastic Metropolis-Hastings

Bayesian methods enjoy high popularity for quantifying uncertainties in complex models. When applied to deep neural networks, we obtain a random network where the distribution of the weights is given by the posterior distribution. In view of large sample sizes and large parameter space dimensions, the design of methods to sample from the posterior is characterized by a tension between numerically feasible and efficient algorithms and approaches which satisfy theoretically justified statistical properties. In this talk we discuss a Bayesian MCMC-based method with a stochastic Metropolis-Hastings as a potential solution. By calculating acceptance probabilities on batches, a stochastic Metropolis-Hastings step saves computational costs, but reduces the effective sample size. We show that this obstacle can be avoided by a simple correction term. We study statistical properties of the resulting stationary distribution of the chain if the corrected stochastic Metropolis-Hastings approach is applied to sample from a Gibbs posterior distribution in a nonparametric regression setting. Focusing on deep neural network regression, we prove a PAC-Bayes oracle inequality which yields optimal contraction rates. The talk is based on joint work with Sebastian Bieringer, Gregor Kasieczka and Maximilian F. Steffen.

Anna Shalova (Eindhoven University)

Title: Singular-limit analysis of gradient descent with noise injection

Abstract:
Noisy gradient descent optimization methods have been empirically shown to find better minimizers compared to deterministic algorithms in various settings. In this work we attempt to give a theoretical explanation of this phenomenon. We study the limiting dynamics of noisy gradient descent systems in the overparameterized regime. In this regime the set of global minimizers of the loss is large, and when initialized in a neighbourhood of this zero-loss set a noisy gradient descent algorithm slowly evolves along this set. We give an explicit characterization of this evolution for the broad class of noisy gradient descent systems. Our results show that the structure of the noise affects not just the form of the limiting process, but also the time scale at which the evolution takes place. We apply our theory to characterize the behavior of various stochastic optimization algorithms including models with dropout, label noise and classical SGD (minibatching) noise.

Samuel Vaiter (Université Côte d’Azur)

Title : Graph Neural Networks on Large Random Graphs

Abstract : Graph Neural Networks (GNNs) are deep architectures defined over graph data that have garnered a lot of attention in recent years. In this talk, I will give some insight on how such architectures behave on random graphs. I will first give non-asymptotic convergence bounds of GNNs toward “continuous” equivalents as the number of nodes grows. Then, I will show their stability to small deformations of the underlying random graph model, a crucial property in traditional CNNs. Finally, I will discuss universality and approximation power with respect to traditional graph tools. This is joint work with Alberto Bietti and Nicolas Keriven.

Yihong Wu (Yale University)

Title: The broken sample problem revisited

Abstract: We revisit the classical broken sample problem: Two sets of iid random vectors $X=\{X_1,\ldots,X_n\}$ and $Y=\{Y_1,\ldots,Y_n\}$ are observed without correspondence. Under the null hypothesis, X and Y are independent. Under the alternative hypothesis, X and Y are correlated in the sense that $(X_i,Y_{\pi(i)})$'s are drawn independently from some bivariate distribution for some unknown permutation π. Originally introduced by DeGroot, Feder, and Goel [AoMS 1971] to model matching records in census data, this problem has gained recently renewed interest due to its applications in data de-anonymization, data integration, and target tracking. Despite extensive research over the past decades, determining the precise statistical thresholds for detection has remained an open problem. In this work, we show that the sharp threshold is given by a spectral and a L_2 condition of the likelihood ratio operator, resolving a conjecture of Bai and Hsing [PTRF 2005] in the positive. Extensions to high dimensions are also discussed. This is based on joint work with Jiaming Xu and Simiao Jiao, both at Duke University.

Jianming Xu (Duke University)

Title: Recent advances on random graph matching problems

Abstract: Random graph matching, the problem of recovering vertex correspondences between two random graphs through their correlated edge connections, is a pivotal challenge with extensive applications. This interdisciplinary problem holds great importance in areas such as network privacy, computational biology, computer vision, and natural language processing, while also raising intricate theoretical questions at the intersection of algorithms, computational complexity, and information theory. Remarkable strides have recently been made in the study of matching correlated Erdős–Rényi graphs. In this talk, the speaker will overview these recent developments, highlight key breakthroughs, and point out important directions for future investigation.