Statistical Aspects of Non-Linear Inverse Problems

Cambridge 17-19 September 2024

Titles and Abstracts

(listed by speaker alphabetically)

Sergios Agapiou (University of Cyprus)

Heavy-tailed Bayesian nonparametric adaptation - direct and inverse problems

We will consider Bayesian nonparametric direct and inverse problems and we will be interested in evaluating the asymptotic performance of the posterior in the infinitely informative data limit, in terms of rates of contraction. We will be especially interested in priors which are adaptive to the smoothness of the unknown function, at least in the direct problem.

In the last decade, certain hierarchical and empirical Bayes procedures based on Gaussian process priors, have been shown to achieve adaptation to spatially homogenous smoothness. However, we have recently shown that Gaussian priors are suboptimal for spatially inhomogeneous unknowns, that is, functions which are smooth in some areas and rough or even discontinuous in other areas of their domain. In contrast, we have shown that (similar) hierarchical and empirical Bayes procedures based on Laplace (series) priors, achieve adaptation to both homogeneously and inhomogeneously smooth functions. All of these procedures involve the tuning of a hyperparameter of the Gaussian or Laplace prior.

After briefly reviewing the above results, we will present a new strategy for adaptation to smoothness based on heavy-tailed priors. We will first illustrate it in the direct setting, showing in particular that adaptive rates of contraction in the minimax sense (up to logarithmic factors) are achieved without tuning of any hyperparameters and for both homogeneously and inhomogeneously smooth unknowns. Then we will show that the adaptation properties of these priors are retained in the diagonal linear inverse problem setting. Finally, we will consider generic forward and inverse problems under a local Lipschitz condition on the forward map, and will study rates of contraction of pseudo-posteriors, that is, posteriors arising from a tempered likelihood. Numerical simulations corroborating the theory will be provided.

The main part of this talk is joint work with Ismaël Castillo.

Giovanni Alberti (University of Genova)

Non-zero constraints in PDE and applications to hybrid inverse problems

The reconstruction in quantitative coupled-physics imaging often requires that the solutions of certain PDEs satisfy some non-zero constraints, such as the absence of critical points or nodal points. I will review several methods that have been employed to construct such solutions, including the Radó–Kneser–Choquet theorem, complex geometrical optics solutions, the use of multiple frequencies, the Runge approximation, the Whitney embedding theorem, and the use of random boundary conditions.

Chiara Amorino (University of Luxembourg)

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

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.

Helmut Bölcskei (ETH Zurich)

Metric-entropy limits on nonlinear dynamical system learning

This talk is concerned with the fundamental limits of nonlinear dynamical system learning from input-output traces. Specifically, we show that recurrent neural networks (RNNs) are capable of learning nonlinear systems that satisfy a Lipschitz property and forget past inputs fast enough in a metric-entropy optimal manner. As the sets of sequence-to-sequence maps realized by the dynamical systems we consider are significantly more massive than function classes generally considered in deep neural network approximation theory, a refined metric-entropy characterization is needed, namely in terms of order, type, and generalized dimension. We compute these quantities for the classes of exponentially-decaying and polynomially-decaying Lipschitz fading-memory systems and show that RNNs can achieve them.

Edoardo Calvello (Caltech)

Transformers for Scientific Machine Learning

Attention mechanisms and their use in transformer architectures have been widely successful at modeling nonlocal correlations in data. Recent interest in applying attention for operator learning motivates a formulation of the methodology in the function space setting. In this talk we outline the construction of an attention mechanism based on [1] in the continuum. We show how this formulation can be leveraged to design transformer neural operators, neural network architectures mapping between infinite-dimensional spaces of functions, and discuss relevant universal approximation theory. By generalizing the “patching” strategy from computer vision to the continuum, we design efficient transformer neural operators which we show to be competitive in cost and accuracy for operator learning tasks involving Darcy flow and Navier-Stokes equations.

[1] Vaswani, Ashish et al. “Attention is All you Need.” *Neural Information Processing Systems* (2017).

Maarten de Hoop (Rice University)

Posterior sampling via score-based diffusion: From finite to infinite dimensions

A Bayesian approach to inverse problems in function spaces has been studied in depth, establishing a foundation that enables further algorithmic and theoretical developments. Recently, score-based diffusion models (SDMs) have shown success as a sampling method in finite-dimensional Bayesian inverse problems. Here, we extend these SDM-based approaches to infinite dimensions while developing a framework for function-space posterior sampling in linear and nonlinear Bayesian inverse problems. For linear problems, we introduce a method to learn posterior distributions using amortized conditional SDMs. Building on prior work by Pidstrigach et al. on learning the unconditional score, we prove the consistency of conditional denoising estimators for infinite-dimensional conditional scores. For learning the score, we employ a neural operator. We also identify the necessary conditions for uniform-in-time estimates of conditional scores for general prior measures, ensuring correct sampling from target conditional distributions. Notably, our analysis reveals that, unlike the unconditional score, the conditional score typically diverges at small times, requiring careful treatment in infinite dimensions. Building on these results in the linear case, we develop a sampling method for nonlinear inverse problems in function spaces that leverages infinite-dimensional SDMs as learning-based priors within a Langevin-type Markov chain Monte Carlo algorithm. Assuming that the forward operator, explicitly in a projection, depends only an arbitrarily finite number of eigenvectors of the relevant trace-class covariance operator for the diffusion prior, we present a convergence analysis with dimension-free bounds that depend on score approximation errors, and are compatible with weighted annealing. We extend the non-asymptotic stationary convergence analysis of Sun et al. from finite to infinite dimensions. Additionally, we provide theoretical guarantees for sampling from posteriors with non-log-concave likelihoods and discuss the robustness of the learned distribution against perturbations in the observations. We showcase our method through both stylized examples and PDE-based inverse problems associated with the acoustic wave equation and the groundwater flow equation. Joint work with L. Baldassari, J. Garnier, A. Siahkoohi and K. Sølna

Zhou Fan (Yale)

Gradient flows for empirical Bayes in high-dimensional linear models

Empirical Bayes provides a powerful approach to learning and adapting to latent structure in data. Theory and algorithms for empirical Bayes have a rich literature for sequence models, but are less understood in settings where latent variables and data interact through more complex designs.

In this work, we study empirical Bayes estimation of an i.i.d. prior in Bayesian linear models, via the nonparametric maximum likelihood estimator (NPMLE). We introduce and study a system of gradient flow equations for optimizing the marginal log-likelihood, jointly over the prior and posterior measures in its Gibbs variational representation using a smoothed reparametrization of the regression coefficients. A diffusion-based implementation yields a Langevin dynamics MCEM algorithm, where the prior law evolves continuously over time to optimize a sequence-model log-likelihood defined by the coordinates of the current Langevin iterate.

We show consistency of the NPMLE under mild conditions, including settings of random sub-Gaussian designs under high-dimensional asymptotics. In high noise, we prove a uniform log-Sobolev inequality for the mixing of Langevin dynamics, for possibly misspecified priors and non-log-concave posteriors. We then establish polynomial-time convergence of the joint gradient flow to a near-NPMLE if the marginal negative log-likelihood is convex in a sub-level set of the initialization. This is joint work with Leying Guan, Yandi Shen, and Yihong Wu.

Matteo Giordano (Turin)

Nonparametric Bayesian intensity estimation for covariate-driven inhomogeneous point processes

The talk will consider nonparametric Bayesian estimation of the intensity function of an inhomogeneous Poisson point process in the important case where the intensity depends on covariates, based on the observation of a single realisation of the point pattern over a large area. It is shown how the presence of covariates allows to borrow information from far away locations in the observation window, enabling consistent inference in the growing domain asymptotics. In particular, minimax-optimal posterior contraction rates under both global and point-wise loss functions are derived. The rates in global loss are obtained under conditions on the prior distribution resembling those in the well established theory of Bayesian nonparametrics, here combined with concentration inequalities for functionals of stationary processes to control certain random covariate-dependent loss functions appearing in the analysis. The local rates are derived with an ad-hoc study that builds on recent advances in the theory of Pólya tree priors, extended to the present multivariate setting with a novel construction that makes use of the random geometry induced by the covariates. Joint work with Alisa Kirichenko and Judith Rousseau.

Barbara Kaltenbacher (University of Klagenfurt)

Convergence guarantees and rates for variational and Newton type methods via range invariance with application in electrical impedance tomography

Range invariance is a property that - like the tangential cone condition - enables a proof of convergence of iterative methods for inverse problems. In contrast to the tangential cone condition it can also be verified for some parameter identification problems in partial differential equations PDEs from boundary measurements, as relevant, e.g., in tomographic applications.

The goal of this talk is to highlight some of these examples of coefficient identification from

boundary observations in elliptic and parabolic PDEs.
In particular, we will also present results on convergence rates for the classical inverse problem of electrical impedance tomography.

Pu-Zhao Kow (National Chengchi University)

Increasing Stability in an inverse boundary value problem and a statistical aspect

Motivated by Abraham and Nickl's recent work about the statistical Calderón problem (Math. Stat. Learn. 2019), we revisit the increasing stability phenomenon in the inverse boundary value problem for the stationary wave equation with a potential using the Bayesian approach. Rather than the well-known and widely used Dirichlet-to-Neumann map, we consider another type of boundary measurements called the impedance-to-Neumann map. Its graph forms a subset of Cauchy data. We will explain the consistency of posterior mean with a contraction rate demonstrating the phenomenon of increasing stability. This talk is prepared based on my recent work with Jenn-Nan Wang (to appear in Taiwanese J. Math.).

Han Lie (University of Potsdam)

Bayesian inference of covariate-parameter relationships for population modelling

An important goal in pharmacology is to tailor drug doses to each patient. To this end, one often uses parametrised ODE initial value problems to model the time evolution of drug concentrations in the body after administration. The parameters in these models often cannot be measured directly in clinical settings, and per-patient data may be too sparse to permit reliable parameter inference for each patient. One approach to this problem is to identify a set of covariates that are clinically measurable, e.g. age and weight, and to specify a covariate-parameter relationship, i.e. a function that maps every admissible covariate vector to a parameter vector. This establishes a nonlinear regression problem, where the i-th covariate X_i is the vector of covariates for the i-th patient, the i-th response Y_i is a vector of blood drug concentrations collected at finitely many times for the i-th patient, and the forward model depends on an unknown covariate-parameter relationship. The task is then to find the most appropriate covariate-parameter relationship from some admissible class. We show how this task can be tackled for a family of parametrised ODEs, by using a framework for Bayesian nonlinear statistical inverse problems developed by Nickl et al., to show posterior contraction and a Bernstein-von Mises result.

Youssef Marzouk (MIT)

Dimension reduction in nonlinear statistical inverse problems

Inverse problems in the Bayesian setting can exhibit many types of "low-dimensional" structure. The posterior distribution might be well approximated as a low-dimensional update of the prior or of some other dominating reference distribution. Alternatively, we might seek low-dimensional projections or summaries of the variables on which we condition. I will survey recent methods for identifying these types of structure, when they are present, and exploiting them in algorithms. A recurrent theme will be the use of gradient-based diagnostics, derived from functional inequalities, that provide upper bounds on the approximation error resulting from dimension reduction.

Andrea Montanari (Stanford)

Posterior Sampling in High Dimensional Linear Regression

I will consider the problem of sampling from the posterior in high-dimensional Bayesian linear regression, when the number of parameters d is of the same order as the number of samples n. I will present two polynomial-time algorithms, the first one based on measure decomposition and the second on diffusion processes.

In both cases (and under different conditions) we prove sampling guarantees

when n> C*d for a suitable constant C.

I will then review other sampling problems that were recently attacked using the diffusion approach, and discuss the fundamental limits of this method.

[Based on joint work with Yuchen Wu, Brice Huang, Huy Tuan Pham, Ahmed El Alaoui, Mark Sellke

Conor Moriarty-Osborne (University of Edinburgh)

Convergence rates of deep Gaussian processes

Gaussian processes have proven to be powerful and flexible tools for various statistical inference and machine learning tasks. However, they can be limited when the underlying datasets exhibit non-stationary or anisotropic properties. Deep Gaussian processes extend the capabilities of standard Gaussian processes by introducing a hierarchical structure, where the outputs of one Gaussian process serve as inputs to another. This hierarchical approach enables deep Gaussian processes to model complex, non-stationary behaviours that standard Gaussian processes may struggle to capture. In this talk, we introduce deep Gaussian processes and explore their use as priors in interpolation and regression tasks. We present results on the convergence rates of deep Gaussian processes in terms of the number of known data points.

Houman Owhadi (Caltech)

Co-discovering graphical structure and functional relationships within data: A Gaussian Process framework for connecting the dots.

Most scientific challenges can be framed into one of the following three levels of complexity of function approximation.

- Type 1: Approximate an unknown function given input/output data.

- Type 2: Consider a collection of variables and functions, some of which are unknown, indexed by the nodes and hyperedges of a hypergraph (a generalized graph where edges can connect more than two vertices). Given partial observations of the variables of the hypergraph (satisfying the functional dependencies imposed by its structure), approximate all the unobserved variables and unknown functions.
- Type 3: Expanding on Type 2, if the hypergraph structure itself is unknown, use partial observations of the variables of the hypergraph to discover its structure and approximate its unknown functions.

Examples of Type 2 problems include solving and learning (possibly stochastic) nonlinear partial differential equations (PDEs), while Type 3 problems encompass learning dependencies between variables in a mechanical system, identifying chemical reaction networks, and determining relationships between genes through a protein-signaling network. Although Gaussian Process (GP) methods are sometimes perceived as a well-founded but old technology limited to Type 1 curve fitting, they can be generalized to an interpretable framework for solving Type 2 and Type 3 problems, all while maintaining the simple and transparent theoretical and computational guarantees of kernel/optimal recovery methods.

Greg Pavliotis (Imperial College London)

Interacting particle systems and their mean field limit: phase transitions, inference and control

In this talk we present recent results on the quantitative study of stochastic interacting particle systems and of their mean field limit. We will start by presenting links between uniform propagation of chaos, the absence of phase transitions and Gaussian fluctuations around the mean field limit. We will present different methodologies for inferring parameters in the mean field (McKean) stochastic differential equation from observations of particle paths. Finally, we show how we can identify stationary states of the mean field PDE and how to steer the dynamics towards a chosen steady state using optimal control methodologies.

Mark Podolskij (University of Luxembourg)

On nonparametric estimation of the interaction function in particle system models

This paper delves into a challenging problem of nonparametric estimation for the interaction function within diffusion-type particle system models. We introduce two estimation methods based upon an empirical risk minimization. Our study encompasses an analysis of the stochastic and approximation errors associated with both procedures, along with an examination of certain minimax lower bounds. In particular, for the first method we show that there is a natural metric under which the corresponding estimation error of the interaction function converges to zero with parametric rate which is minimax optimal. This result is rather surprising given the complexity of the underlying estimation problem and rather large class of interaction functions for which the above parametric rate holds.

Kolyan Ray (Imperial College London)

Bayesian nonparametric inference in a McKean-Vlasov model

We study nonparametric estimation of the interaction term in a McKean-Vlasov model where noisy observations are drawn from the nonlinear parabolic PDE arising in the mean-field limit as the number particles grows to infinity. In this model, the long-time invariant state can be uninformative about the interaction potential. We therefore show that under certain regularity conditions on the initial state, the short-time behaviour of this system contains sufficient information to consistently recover the interaction potential using Gaussian process priors. This involves establishing a stability-type estimate for this PDE to solve the resulting inverse problem.

This is joint work with Richard Nickl and Greg Pavliotis.

Markus Reiss (Humboldt University)

Statistics for SPDEs

Paul Rosa (Oxford University)

Nonparametric regression on random geometric graphs sampled from submanifolds

We consider the nonparametric regression problem when the covariates are located on an unknown smooth compact submanifold of a Euclidean space. Under defining a random geometric graph structure over the covariates we analyze the asymptotic frequentist behaviour of the posterior distribution arising from Bayesian priors designed through random basis expansion in the graph Laplacian eigenbasis. Under Holder smoothness assumption on the regression function and the density of the covariates over the submanifold, we prove that the posterior contraction rates of such methods are minimax optimal (up to logarithmic factors) for any positive smoothness index.

Otmar Scherzer (University of Vienna)

Numerical Linear Algebra Networks for Solving Linear Inverse Problems

We consider solving a probably ill-conditioned linear operator equation, where the operator is not modelled but specified via training pairs of the input-output relation of the operator. The proposed method is motivated from de-and encoder networks strategies for solving nonlinear inverse problems. The linear case is simpler and by showing synergies we want to find more insight in the structure of coding networks.

Christoph Schwab (ETH Zurich)

Multilevel approximation of Gaussian random fields

Centered Gaussian random fields (GRFs) indexed by compacta
as e.g. compact orientable manifolds M are determined by their covariance operators.

We consider the numerical analysis of sample-wise, compressive multi-level wavelet-Galerkin approximations of centered GRFs given as variational solutions to coloring operator equations driven by spatial white noise, with pseudodifferential covariance operator being elliptic, self-adjoint and positive from the Hörmander class.

For pathwise approximations with p parameters, tapered covariance or precision matrices have O(p) nonzero entries, can be optimally diagonally preconditioned,
and allow O(p) path simulation, covariance estimation and kriging of GRFs.

Joint work with Helmut Harbrecht (Uni Basel),
Kristin Kirchner (TU Delft), and Lukas Herrmann (RICAM, Linz).

Maximilian Siebel (Heidelberg)

Convergence Rates for the Maximum A Posteriori Estimator in PDE-Regression Models with Random Design

In this ongoing work, we consider the statistical inverse problem of recovering a parameter $\theta\in H^\alpha$ from data arising from the Gaussian regression problem

\begin{equation*}

Y = \mathcal{G}(\theta)(Z)+\sigma\varepsilon

\end{equation*}

with nonlinear forward map $\mathcal{G}:\mathbb{L}^2\to\mathbb{L}^2$ and random design points Z. The estimation strategy is based on a least squares approach under H^α-constraints. Under Lipschitz-type assumptions on the forward map $\mathcal{G}$, we establish the existence of a least squares estimator $\hat{\theta}$ as a maximizer for some given functional. We state a general concentration result, which is used to prove consistency and upper bounds for the prediction error. The corresponding rates of convergence reflect not only the smoothness of the parameter of interest but also the ill-posedness of the underlying inverse problem. We apply the general model to the Darcy problem, where the recovery of an unknown coefficient function of a PDE is of interest. For this example, we also provide corresponding rates of convergence for the prediction and estimation errors. Additionally, we briefly discuss the applicability of the general model to other problems.

Michael Sørensen, (University of Copenhagen)

Diffusion processes on the torus - models of time series of angular data

Two classes of diffusion processes on the multivariate torus with related statistical methodology are presented. The aim is to model time series of angular data. The diffusion processes are ergodic and time-reversible and can be constructed for any pre-specified stationary distribution on the torus. Applications to the evolution of proteins and to ants' movement are briefly presented.

For the class of Langevin diffusions, approximations to the likelihood function are presented and compared. We also present a class of diffusion models with explicit transition probability densities, which enables exact likelihood inference. We consider asymptotic likelihood theory and easy exact diffusion bridge simulation for the latter class. A class of circular jump processes with similar properties is proposed too.

Co-author: Eduardo García-Portugués, Department of Statistics, Carlos III University of Madri

Vladimir Spokoiny (Humboldt University of Berlin)

Statistical inference for nonlinear inverse problems»

The talk discusses a new approach to statistical analysis for a large class of statistical models including nonlinear inverse problems. The main idea behind the method is to extend the parameter space and to replace the structural equation with a structural penalty.

The focus is on finite sample expansions for the profile MLE in the extended model which enable us to obtain sharp risk bounds and to study the asymptotic properties of the estimator. We provide a few examples illustrating the approach for elliptic PDEs.

Bjorn Sprungk (TU Freiberg)

Noise-level robust sampling and Bayesian inference on the sphere

We consider two topics in this talk: the first is related to sampling from concentrated posterior distributions arising in Bayesian inference with informative data. Although a desirable situation from an inference perspective concentrated posteriors pose a computational challenge for many sampling algorithms. We present results regarding Markov chain Monte Carlo methods based on the Laplace approximation which show a statistical efficiency independent of the concentration of the posterior under suitable assumptions.
The second topic again considers the construction of Markov chain Monte Carlo algorithms but this time for dimension-independent sampling from posterior measures defined on a high-dimensional sphere as occurring in Bayesian density estimation. Both topics are related by the concept of pushforward Markov kernels.

Bernard Stankewitz (University of Milano Bocconi)

Contraction rates for conjugate gradient and Lanczos approximate posteriors in Gaussian process regression

Due to their flexibility and theoretical tractability Gaussian process (GP) regression models have become a central topic in modern statistics and machine learning.

While the true posterior in these models is given explicitly, numerical evaluations depend on the inversion of the augmented kernel matrix $K + \sigma^2 I$, which requires up to $O(n^3)$ operations.
For large sample sizes n, which are typically given in modern applications, this is computationally infeasible and necessitates the use of an approximate version of the posterior. Although such methods are widely used in practice, they typically have limited theoretical underpinning.

In this context, we analyze a class of recently proposed approximation algorithms from the field of Probabilistic numerics.
They can be interpreted in terms of Lanczos approximate eigenvectors of the kernel matrix or a conjugate gradient approximation of the posterior mean, which are particularly advantageous in truly large scale applications, as they are only based on matrix vector multiplications amenable to the GPU acceleration of modern software frameworks.

We combine result from the numerical analysis literature with state of the art concentration results for spectra of kernel matrices to obtain minimax contraction rates.

Andrew Stuart (Caltech)

The Mean-Field Ensemble Kalman Filter

Ensemble Kalman filters constitute a methodology approximating aspects of the filtering distribution in partially observed and noisy dynamical systems. They are widely adopted in the geophysical sciences, underpinning weather forecasting for example, and are starting to be used throughout the sciences and engineering; furthermore, they have been adapted to function as a general-purpose tool for parametric inference. The strength of these methods stems from their ability to operate using complex models as a black box, together with their natural adaptation to high performance computers. In this talk we introduce theory which elucidates conditions under which this widely adopted methodology provides accurate model predictions and uncertainties for discrete time filtering. The theory rests on a mean-field formulation of the methodology and an error analysis controlling differences between probability measure propagation under the mean-field model and under the true filtering distribution.

The mean-field formulation is based on joint work with Edoardo Calvello (Caltech) and Sebastian Reich (Potsdam).

The error analysis is based on joint work with Jose Carrillo (Oxford), Franca Hoffmann (Caltech) and Urbain Vaes (Paris) and on joint work with Edoardo Calvello (Caltech), Pierre Monmarche and Urbain Vaes (Paris).

Botond Szabo (University of Bocconi)

Linear methods for non-linear inverse problems

We consider recovering an unknown function f from a noisy observation of the solution

u_f to a partial differential equation of the type $\mathcal{L} u_f=c(f,u_f)$

for a differential operator $\mathcal{L}$, and invertible function c, i.e. $f=e(\mathcal{L}u_f)$.

Examples include amongst others the time-independent Schr\"odinger equation $\frac{1}{2}\Delta u_f = u_ff$ and the heat equation with absorption term $\frac{d u_f}{dt}-\frac{1}{2}\Delta u_f=f$. We transform this problem into the linear inverse problem of recovering $\mathcal{L}u_f$ under Dirichlet boundary condition, and show that Bayesian methods (with priors placed either on u_f or $\mathcal{L}u_f$) for this problem may yield optimal recovery rates not only

for u_f, but also for f. We also derive frequentist coverage guarantees for the corresponding Bayesian credible sets. Adaptive

priors are shown to yield adaptive contraction rates for f, thus eliminating the need to know the smoothness of this function. The results are illustrated by several numerical analysis on synthetic data sets. Joint work with Aad van der Vaart (Delft) and Geerten Koers (Delft)

Edriss S. Titi (University of Cambridge Texas A&M University and Weizmann Institute of Science)

Rigorous Analysis and Numerical Implementation of Nudging Data Assimilation Algorithms

In this talk, we will introduce downscaling data assimilation algorithms for weather and climate prediction based on discrete coarse spatial scale measurements of the state variables (or only part of them, depending on the underlying model). The algorithm is based on linear nudging of **the coarse spatial scales** in the algorithm's solution toward the coarse spatial scales corresponding to the observed measurements of the unknown reference solution. The algorithm's solution can be initialized arbitrary and is shown to converge at an exponential rate toward the exact unknown reference solution. This indicates that the dynamics of the algorithm is globally stable (not chaotic) unlike the dynamics of the model that governs the unknown reference solution. Capitalizing on this fact, we will also demonstrate uniform in time error estimates of the numerical discretization of these algorithms, which makes them reliable upon implementation computationally. Furthermore, we will also present a recent improvement of this algorithm by employing nonlinear nudging, which yields super exponential convergence rate toward the unknown exact reference solution.

Sven Wang (Humboldt University)

Likelihood-based methods for low frequency diffusion data

We consider the problem of nonparametric inference in multi-dimensional diffusion models from low-frequency data. Due to the computational intractability of the likelihood, implementation of likelihood-based procedures in such settings is a notoriously difficult task. Exploiting the underlying (parabolic) PDE structure of the transition densities, we derive computable formulas for the likelihood function and its gradients. We then construct a Metropolis-Hastings Crank-Nicolson-type algorithm for Bayesian inference with Gaussian priors, as well as gradient-based methods for computing the MLE and Langevin-type MCMC. The performance of the algorithms is illustrated via numerical experiments.

Jakob Zech (Heidelberg University)

Statistical Learning Theory for Neural Operators

In this talk, we discuss convergence rates for neural network-based operator surrogates, which approximate smooth maps between infinite-dimensional Hilbert spaces. Such surrogates have a wide range of applications and can be used in uncertainty quantification and parameter estimation problems in fields such as classical mechanics, fluid mechanics, electrodynamics, earth sciences etc. Here, the operator input represents the problem configuration and models initial conditions, material properties, forcing terms, and/or the domain of a partial differential equation (PDE) describing the underlying physics. The operator output is the corresponding PDE solution. Our analysis demonstrates that, under suitable smoothness assumptions, the empirical risk minimizer for specific neural network architectures can overcome the curse of dimensionality both in terms of required network parameters and the input-output pairs needed for training.

Yichen Zhu (Bocconi University)

Vecchia Gaussian Processes: Probabilistic properties and Bayesian Nonparametrics

Gaussian Processes are widely used to model spatial dependency in geostatistical data, yet the exact computation suffers an intractable time complexity of $O(n^3)$. Vecchia approximation allows scalable Bayesian inference of Gaussian processes in $O(n)$ time by introducing sparsity in the spatial dependency structure that is characterized by a sparse directed acyclic graph (DAG). Despite the popularity in practice, little is understood about the Vecchia Gaussian processes themselves, let alone their theoretical guarantees when employed in regression models. In this paper, we systematically study the probabilistic properties of Vecchia Gaussian processes when the mother Gaussian process is a Mat\'{e}rn process. Under minimal regularity conditions and appropriate selection of the DAG, the Vecchia Gaussian process retains many desirable properties of the mother Gaussian process.

These probabilistic properties further allow us to develop Bayesian nonparametric theory for the Vecchia Gaussian process, where minimax optimality is achieved by optimally tuned Gaussian processes via either oracle rescaling or hierarchical Bayesian methods.